

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:27:13 ; Search time 14 Seconds
(Without alignments)
1653.128 Million cell updates/sec

Title: US-10-037-311a-1
Perfect score: 3004
Sequence: 1 MDQNSYRRSSPIRTTGS.....GTLVPHVRCEDISWGLKLV 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3004	100.0	558	1	095wh5 arabidopsis
2	1776	59.1	565	1	09m5q1 pisum sativ
3	1706.5	56.8	526	1	09x181 arabidopsis
4	1679	55.9	539	1	081053 arabidopsis
5	1613.5	53.7	500	1	09x178 arabidopsis
6	1583	52.7	537	1	09x180 arabidopsis
7	1528	50.9	503	1	09x180 arabidopsis
8	1492	49.7	533	1	09x180 arabidopsis
9	1411.5	47.0	440	1	09x180 arabidopsis
10	1302	43.3	435	1	09x180 arabidopsis
11	1113.5	37.1	493	1	09x180 arabidopsis
12	110.5	3.7	322	1	09x180 arabidopsis
13	97	3.2	594	1	09x180 arabidopsis
14	97	3.2	1224	1	09x180 arabidopsis
15	96	3.2	594	1	09x180 arabidopsis
16	94	3.1	1333	1	09x180 arabidopsis
17	94	3.1	1376	1	09x180 arabidopsis
18	94	3.1	1376	1	09x180 arabidopsis
19	93.5	3.1	1376	1	09x180 arabidopsis
20	93	3.1	1376	1	09x180 arabidopsis
21	93	3.1	1376	1	09x180 arabidopsis
22	92.5	3.1	1376	1	09x180 arabidopsis
23	92	3.1	1376	1	09x180 arabidopsis
24	92	3.1	1376	1	09x180 arabidopsis
25	92	3.1	1376	1	09x180 arabidopsis
26	91	3.0	1376	1	09x180 arabidopsis
27	90.5	3.0	1376	1	09x180 arabidopsis
28	90.5	3.0	1376	1	09x180 arabidopsis
29	90.5	3.0	1376	1	09x180 arabidopsis
30	90.5	3.0	1376	1	09x180 arabidopsis
31	90.5	3.0	1376	1	09x180 arabidopsis
32	90	3.0	1376	1	09x180 arabidopsis
33	90	3.0	1376	1	09x180 arabidopsis

34	90	3.0	3080	1	POLG_ZYMW
35	90	3.0	3746	1	ACV5_PENCH
36	90	3.0	3791	1	ACVT_PENCH
37	89.5	3.0	1286	1	PATC_DROME
38	89.5	3.0	1460	1	NI59_YEAST
39	89	3.0	369	1	GAL7_SCHPO
40	89	3.0	3083	1	POLG_ZYMW
41	88.5	2.9	1693	1	RIP2_MOUSE
42	88.5	2.9	2116	1	RRPL_SYNY
43	88	2.9	936	1	RPOP_CLAPU
44	88	2.9	957	1	KE5C_HUMAN
45	88	2.9	1324	1	VGL2_CVMA5

ALIGNMENTS

RESULT 1
FUT1_ARATH
ID FUT1_ARATH STANDARD: PRT: 558 AA.
AC 095WH5: 081052;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galactoside 2-L-fucosyltransferase (EC 2.4.1.69) (Xyloglucan alpha-
DE (1,2)-fucosyltransferase (AtrFUT1).
GN FUT1 OR FRT1 OR AT2G03220 OR T18E12.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=cv. Columbia.
RA MEDLINE=99301928; PubMed=10373113;
RA Perrin R.M., DeRoche A.E., Bar-Peled M., Zeng W., Norambuena L.,
RA Orellana A., Ralshel N.V., Keegstra K.;
RT "Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall
RT biosynthesis."
RT Science 284:1976-1979(1999).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia.
RA MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umeyam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Newman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
[3]
RP FUNCTION, AND TISSUE SPECIFICITY.
RA MEDLINE=21608393; PubMed=11743104;
RA Sarria R., Wagner T.A., O'Neill M.A., Falk A., Wilkerson C.G.,
RA Keegstra K., Ralshel N.V.;
RT "Characterization of a family of Arabidopsis genes related to
RT xyloglucan fucosyltransferase."
RL Plant Physiol. 127:1595-1606(2001).
CC -i- FUNCTION: involved in cell wall biosynthesis. Adds the terminal
CC fucosyl residue on xyloglucan side chains.
CC -i- CATALYTIC ACTIVITY: GDP-L-fucose + beta-D-galactosyl-R = GDP +
CC alpha-L-fucosyl-1,2-beta-D-galactosyl-R.
CC -i- PATHWAY: Glycosylation.
CC -i- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC -i- TISSUE SPECIFICITY: Expressed in roots, stems, leaves, flowers,
CC siliques and seedlings.

```
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF154111; AAD41092.1; -
CC EMBL: AC005313; AAC34480.1; -
CC InterPro: IPR004938; XG_Ftase; 1.
CC Pfam: PF03254; XG_Ftase; 1.
CC Transferase: Glycosyltransferase; Transmembrane; Glycoprotein;
CC Signal-anchor; Golgi stack; Cell wall.
CC DOMAIN 1 43 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 44 64 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC DOMAIN 65 558 LUMENAL, CATALYTIC (POTENTIAL).
CC CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 270 270 V -> F (IN REF. 1).
CC SEQUENCE 558 AA; 63452 MW; 1BF0E372231A8ADF CRC64;

Query Match 100.0%; Score 3004; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 3.5e-234;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOONSRRRRSIRRTTGGSKSVNSSELLQMKYLSGGTKMLRTTTCIVSVLVAESM 60
DB 1 MOONSRRRRSIRRTTGGSKSVNSSELLQMKYLSGGTKMLRTTTCIVSVLVAESM 60
QY 61 IFHQHPSDSNRIMGFAEARVLDAGVFPNTNNSDKLGGILASGEDSCSRQSVHY 120
DB 61 IFHQHPSDSNRIMGFAEARVLDAGVFPNTNNSDKLGGILASGEDSCSRQSVHY 120
QY 121 RKPSPYKSSYISIKLRNTEKLRKCGPTESYKALQOLDQEHIDGDECKYVWISFS 180
DB 121 RKPSPYKSSYISIKLRNTEKLRKCGPTESYKALQOLDQEHIDGDECKYVWISFS 180
QY 121 RKPSPYKSSYISIKLRNTEKLRKCGPTESYKALQOLDQEHIDGDECKYVWISFS 180
DB 121 RKPSPYKSSYISIKLRNTEKLRKCGPTESYKALQOLDQEHIDGDECKYVWISFS 180
QY 181 GIGNNLSIASVFLYALLTDRLVLDNRKMDMLCCEPFLGMSWLLPDLFPMTDOFDGLN 240
DB 181 GIGNNLSIASVFLYALLTDRLVLDNRKMDMLCCEPFLGMSWLLPDLFPMTDOFDGLN 240
QY 241 QESSRCYGYMKNQYIDTEGLTSLHLYLHVHDYGDHDKMFCGEGQPTLGKVPMLIVKTD 300
DB 241 QESSRCYGYMKNQYIDTEGLTSLHLYLHVHDYGDHDKMFCGEGQPTLGKVPMLIVKTD 300
QY 301 NYFVPSMLIGFDELAKLPQKATVHHILGRYLFHPTNOVWGLVTRYEAYISHADEK 360
DB 301 NYFVPSMLIGFDELAKLPQKATVHHILGRYLFHPTNOVWGLVTRYEAYISHADEK 360
QY 301 NYFVPSMLIGFDELAKLPQKATVHHILGRYLFHPTNOVWGLVTRYEAYISHADEK 360
DB 301 NYFVPSMLIGFDELAKLPQKATVHHILGRYLFHPTNOVWGLVTRYEAYISHADEK 360
QY 361 IGIOVRVDEDEPPRQOHMDIOISCTQEKLLPEYDVLVESRHVNTPKHAAVLVTSINA 420
DB 361 IGIOVRVDEDEPPRQOHMDIOISCTQEKLLPEYDVLVESRHVNTPKHAAVLVTSINA 420
QY 421 GYAENIKSMWEYPTSTGEIIGVHQPQSEGYQOTEKKHNGKALAEMLLSLTDNLVISA 480
DB 421 GYAENIKSMWEYPTSTGEIIGVHQPQSEGYQOTEKKHNGKALAEMLLSLTDNLVISA 480
QY 481 WSTGYYVAQGIIGLKPWLLPPEKTTTDPSCGRAMSEPCFHSPPDYDCAKGTIDGT 540
DB 481 WSTGYYVAQGIIGLKPWLLPPEKTTTDPSCGRAMSEPCFHSPPDYDCAKGTIDGT 540
QY 541 LVPHVRHCEDISWGLKIV 558
DB 541 LVPHVRHCEDISWGLKIV 558
DB 541 LVPHVRHCEDISWGLKIV 558

RESULT 2
FUT1_PEA STANDARD; PRT; 565 AA.
ID FUT1_PEA
```

```
AC Q9W501;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Galactoside 2-L-fucosyltransferase (EC 2.4.1.69) (Xyloglucan alpha-
DE (1,2)-fucosyltransferase (PsfT1)).
GN PSF1.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=cv Alaska;
RX MEDLINE=20270214; PubMed=10747946;
RA Falk A., Bar-Peled M., Derocher A.E., Zeng W., Perrin R.M.,
RA Wilkerson C., Raikhel N.V., Keegstra K.;
RT "Biochemical characterization and molecular cloning of an
RT alpha-1,2-fucosyltransferase that catalyzes the last step of cell wall
RT xyloglucan biosynthesis in pea."
RL J. Biol. Chem. 275:15082-15089(2000).
CC -1- FUNCTION: Involved in cell wall biosynthesis. Adds the terminal
CC fucosyl residue on xyloglucan side chains.
CC -1- CATALYTIC ACTIVITY: GDP-L-fucose + beta-D-galactosyl-R = GDP +
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF223643; AA62886.1; -
CC InterPro: IPR004938; XG_Ftase; 1.
CC Pfam: PF03254; XG_Ftase; 1.
CC Transferase: Glycosyltransferase; Transmembrane; Glycoprotein;
CC Signal-anchor; Golgi stack; Cell wall.
CC DOMAIN 1 43 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 44 64 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC DOMAIN 65 565 LUMENAL, CATALYTIC (POTENTIAL).
CC CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 509 509 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 565 AA; 64029 MW; 5E6AAABED20594B CRC64;

Query Match 59.1%; Score 1776; DB 1; Length 565;
Best Local Similarity 52.6%; Pred. No. 2.8e-135;
Matches 330; Conservative 69; Mismatches 116; Indels 12; Gaps 4;

QY 39 MKLRTTTCIVSVLVAESMIFHQHPSD---SNRIMGFAEARVLDAGVFPNTNNSD 95
DB 42 MRVAFVVSFMFLSVFLSVLDRPPSDAALSTTLTFLQNLGSLGSDSDFSVELLN 100
QY 96 KILGGLASGEDSCSRQSVHYRKPSPYKSSYISIKLRNTEKLRKCGPTESYK 155
DB 101 KILGGLADGFDEKSCSRQSAIFGKLSGKSSYISIRKRYEAKHKCCGPTESYK 160
QY 156 ALKQLDQEHIDGDECKYVWISFSFGNRIILASVFLYALLTDRLVLDNRKMDMLF 215
DB 161 TYKELSGQFSSEYDCKYVWISFSGLGNRIILTVSAFLYALLTDRLVLDNRKMDMLF 220
QY 216 CEPFLGMSWLLPDLFPMTDOFDGLNDESSRCYGYMKNQYIDTEGLTSLHLYLVDYGD 275
DB 221 CEPFPDASWFWPPDFPLNSHLNNEFNSNCHGKIITKTSITNSTVPSFYVLIHLADYDD 280
```


us-10-037-311a-1.rsp

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae,
OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buchter E., Chanson A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gull J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huzar L.
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetskaia I., Kuttz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lens C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marshall A.,
RA Miltiescher J., Miranda M., Nguyen M., Niemann W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utechtack T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Winter J.C., Davis R.W.;
RN "Sequence and analysis of chromosome 1 of the plant Arabidopsis
UT

RL Nature 408:816-820(2000).
RN [2]
RP IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY

RA Sarria R., Wagner T.A., O'Neill M.A., Faik A., Wilkerson C.G.,
RA Keegstra K., Ralshel N.V.;

RT xyloglucan fucosyltransferase1.";
RL Plant Physiol. 127:1595-1606(2001).

CC act as a fucosyltransferase.
CC -1- PATHWAY: Glycosylation.

CC form in trans cisternae of Golgi (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed in roots and flowers.

-1- CAUTION: Ref.1 (Aar79408) sequence differs from that shown due to erroneous gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collabora
CC between the Swiss Institute of Bioinformatics and the EMBL outstat

CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for comm-

```

CC      or send an email to license@sb-slb.ch.
CC
CC
CC      -----
CC      FROM: license@sb-slb.ch
CC

```

DR EMBL: AC068119.1; AAF/9408.1; ALT_SEQ.
DR InterPro: IPR004938; XG_FTase.
DR RefSeq: NP03354.1; WC_FTase.
DR

KW translocase; glycosyltransferase; transmembrane; glycoprotein;
KW signal-anchor; Golgi stack; cell wall.
ET DOMAIN 1 CYTODERMIC (PLASMA) MEMBRANE

DOMAIN	442	537
FT	21	41
FT	1KANSMEM	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT		(POTENTIAL).
FT		TIEMNAL CATAIVTIC /DOTENTIAL\

[illegible]

Query Match	52.7%	Score 1583	DB 1	length 537
SEQUENCE	CGT	01191	MM	4200473004334626
				CINCO4

Matches 317; Conservative 71; Mismatches 122; Indels 26; Gaps

[illegible]


```

AC 095Jp4; (Rel. 41, Created)
AD 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable fucosyltransferase 5 (EC 2.4.1.-) (AtFUT5).
GN FURS OR AT2G15370 OR F26H6.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A..
RX MEDLINE=21608393; PubMed=11743104;
RA Satria R., Wagner T.A., O'Neill M.A., Paik A., Wilkerson C.G.,
RA Keegstra K., Raikhel N.V.;
RT "Characterization of a family of Arabidopsis genes related to
RT xyloglucan fucosyltransferase."
RL Plant Physiol. 127:1595-1606(2001).
RN [2]
RP SEQUENCE FROM N.A..
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell K.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.-L.,
RA Mofat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Mierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
CC -I- FUNCTION: May be involved in cell wall biosynthesis. May
CC act as a fucosyltransferase.
CC -I- PATHWAY: Glycosylation.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC -I- TISSUE SPECIFICITY: Expressed in roots, leaves, flowers and
CC siliques.
CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AF417475; AAL50624.1; ALT_INIT.
CC DR EMBL; AC006920; AAD22287.1; -.
CC DR InterPro; IPR004938; XC_FTase.
CC Pfam; PF03254; XG_FTase; 1.
CC KW Transferase; Glycosyltransferase; Transmembrane; glycoprotein;
CC Signal-anchor; Golgi stack; Cell wall.
CC FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 14 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC LUMENAL, CATALYTIC (POTENTIAL).
CC FT DOMAIN 35 533
CC CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 533 AA; 61173 MW; 63CE63AFAD56E43 CRC64;
Query Match 49.7%; Score 1492; DB 1; Length 533;
Best Local Similarity 54.6%; Pred. no. 1,9e-112;
Matches 295; Conservative 90; Mismatches 133; Indels 22; Gaps 9;

```

Db	1	MYKFOISG - KIVKTLGLKKAKVLLIAVSGSLF-----ILSTSNFNKKLDDATT	49
Qy	86	FPNV - ININSDKLLGGSLSEDEDSCLRYOSVH - YRKSPYKESSYLSTKLRYEKL	142
Db	50	KVDIKETFEKRPDKLIGGLLTAADPESSCLSRHKYFLYRKRPSPYKRSYLVLSKLSREWL	109
Qy	143	HKRCGPGETESYKAKLKOLDQEH - DGDGECKYVWVIFSGLGNRIILSLASVFLYALLTR	201
Db	110	HKRCGPDEYKKEAIETKLSRDASSESNQECRYIWMVAGYGLGNRIILTLASVFLYALLTER	169
Qy	202	VLLDRGCDMDLCEPFFLGSLWLLPLDFFMTQFD--GINQSSSCYCYMKQNYIDE	259
Db	170	IILVDNRKVDSDLLCEPFPGLSWLLPLDFFMLVITYAMGYNKPEPCYGMSEKHSINST	229
Qy	260	GLTSLHLYHLVDYGDHDKMFCFGDDOTFGKQWMLVTKDNYEVSLEMLPECFDELK	319
Db	230	SIRPHILYHNLHSDRSDSKLFYCKQKDSLIDKPYMLIVQANYVFSVLFNPFQIETLK	289
Qy	320	LEPKATVFNHLLGRLYLFHPTNQWGLVTRYEAVLSHADEKIGQYRVDEDPQPHWM	379
Db	290	LEPKETVEFHLLARYLFHPNEVMDVWTDYHMLSKADERLGIQLRGVKPDGRKXK	349
Qy	360	DOISSCTQEKLEPVDLTVERSHVN - TRKRAVLVTSLNAGYAENLSMWEYPTSTG	438
Db	350	DOVISCQREKLEFPATPEESKKNISKTRKLSVLVASLPEFSGTLNMFESKRPSSTG	409
Qy	439	EIIHGPOSGCYOQTEKKHNGKALAEMLLSTDLVLSAMSTGYVAGQGLKPMI	498
Db	410	EIVEYQPSGGRVQOTDKSHIDKALAEMLLSTDLNIVTSANSTGYVSYSLGGLKPMI	469
Qy	499	LYREPNRTTDPSPSGRAMSMPCFSPFPYDCAKTKTIDGTLPVPHRCEDISW - GLKL	557
Db	470	LYQPNFTTTPMPCEVRKSMPCYLTPPSSHGCEADMGNTSGKITLPVRHCEDLITYGLKL	529
RESULT 9			
FU01_	FU01_	ARATH	STANDARD; PRT; 440 AA.
AC	Q95JP6		
DT	15-JUN-2002	(Rel. 41, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Putative fucosyltransferase 10 (EC 2.4.1.-) (AtFU10) (Fragment).		
GN	FU10 OR AT2615350 OR F26H6.13.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryote: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;		
OC	Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;		
OC	eucots II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Columbia;		
RC	MEDLINE=20083487; PubMed=10617197;		
RA	Liu X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,		
RA	Fuji C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,		
RA	Boffai K.S., Keichum K.A., Lee J.J., Ronning C.M., Koo H.L.,		
RA	Muehl K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,		
RA	Tallon L.J., Gill J.E., Adams M.D., Carreira A.J., Cready T.H.,		
RA	Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,		
RA	Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,		
RA	Venter J.C.;		
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis		
RT	thaliana.";		
RL	Nature 402:761-768(1999).		
RN	[2]		
RP	IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY		
RP	MEDLINE=21608393; PubMed=11743104;		
RA	Sarría R., Wagner T.A., O'Neill M.A., Falk A., Wilkerson C.G.,		
RA	Keegstra K., Raskhel N.V.,		
RT	"Characterization of a family of Arabidopsis genes related to		
RT	xyloglucan fucosyltransferase1.";		
RL	Plant Physiol. 127:1595-1606(2001).		

```

CC -1- FUNCTION: May be involved in cell wall biosynthesis. May
CC act as a fucosyltransferase.
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Golgi (potential).
CC -1- TISSUE SPECIFICITY: Expressed in root, leaves, stems and
CC seedlings.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -1- CAUTION: AtFUT10 (as currently annotated) is lacking the first
CC exon and intron. It is therefore incomplete and we awaits cDNA
CC confirmation of the N-terminal extension.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC006920; AAD22289.1; -.
CC InterPro: IPR004938; XG_FTase.
CC Pfam: PF03254; XG_FTase; 1.
CC DR Hypothetical protein; Transferase; Glycosyltransferase; Glycoprotein;
CC Golgi stack; Cell wall.
CC KW NON-TER 1
CC FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 440 AA; 50837 MW; 2469BD38934080 CRC64.

Query Match 47.0%; Score 1411.5; DB 1; Length 440;
Best Local Similarity 59.5%; Pred. No. 4.4e-106;
Matches 261; Conservative 75; Mismatches 90; Indels 13; Gaps 5;

OY 128 PSSYLISKRNNEKLRKRCGPETSEYKAKLKOLDQEH--DGDGECKYVWISFSGIGLN 184
DB 2 PSEYLVSELRSEYEMLRKRGPTKATEKTLSDREYVASESNGECRYIWLARDGICLN 61
OY 185 RLTLASVFLVALLDRLVLDVRGKMDLDFCEPLGSMWLPDLPMPMD-QDFGLDNE 242
DB 62 RLTLASVFLVALLDRLVLDVRGKMDLDFCEPLGSMWLPDLPMPMD-QDFGLDNE 121
OY 243 SSRICGVYMKNOVIDEGLSLHLYLHLVHDYDHDHMEFCEGDQFTIGKVPWLYIKTDNY 302
DB 122 YRCYGTMLNENIAINSTSIPRLYLHNHDSRDSKLFECQKDSFIDKVPWLIQTNAV 181
OY 303 FVPSLWLVGDFDELNKLPKOKATVFHHGRYLFHPNQVWGLVTRYEATYSHADEKIG 362
DB 182 FVPSLWLVGDFDELNKLPKOKATVFHHGRYLFHPNQVWGLVTRYEATYSHADEKIG 241
OY 363 IOVRVDEDDPGFPOHVMDISSCTOKELLPVDFLVERSRHVN--TPKRAVAVTSIN 419
DB 242 IOVRVDEDDPGFPOHVMDISSCTOKELLPVDFLVERSRHVN--TPKRAVAVTSIN 297
OY 420 AGYAENLMSMYEPTSTGEITIGVHQSOGEGYQTEKKMHNSKALAEWYLLSTDNLVTS 479
DB 298 PERYSVNLTMFLARSSSTGEILEVYQPSAEVYQDQDKKSHQKALAEWYLLSTDNLVTS 357
OY 480 AMSTGCVYVAGGGLKPMPLVREKRTTPDPSGGRAMSECFHSPPRYDDCAKAKIDTG 539
DB 356 GMSITGVYYSISGLGKPMPLVLOPVNFTTPNPPCVNRKSMDECPYHPPSHGCEADWGTNSG 417
OY 540 TLVPHVRHCEIDSM-GLKL 557
DB 418 KILPVRHCEIDSMYGLKL 436

RESULT 10
FUT9_ARATH STANDARD; PRT; 435 AA.
ID FUT9_ARATH
AC Q9X177;

```

```

DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable fucosyltransferase 9 (EC 2.4.1.-) (AtFUT9).
GN FUT9 OR AT1G14110 OR F7A19.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia.
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Paim C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Talon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uteckack T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
RN [2]
RP IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY.
RX MEDLINE=21608393; PubMed=11743104;
RA Sarría R., Wagner T.A., O'Neill M.A., Falk A., Wilkerson C.G.,
RA Keegstra K., Raikhel N.V.;
RT "Characterization of a family of Arabidopsis genes related to
RT xyloglucan fucosyltransferase1."
RL Plant Physiol. 127:1595-1606(2001).
CC -1- FUNCTION: May be involved in cell wall biosynthesis. May
CC act as a fucosyltransferase.
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed in leaves and stems.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC007576; AAD39295.1; -.
CC InterPro: IPR004938; XG_FTase.
CC Pfam: PF03254; XG_FTase; 1.
CC DR Transferrase; Glycosyltransferase; Transmembrane; Glycoprotein;
CC Signal-anchor; Golgi stack; Cell wall.
CC KW TRANSMEM 1
CC FT DOMAIN 22 435
CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 435 AA; 49976 MW; DE26E49EDB69629 CRC64;

Query Match 43.3%; Score 1302; DB 1; Length 435;
Best Local Similarity 51.0%; Pred. No. 2.9e-97;
Matches 268; Conservative 57; Mismatches 97; Indels 104; Gaps 9;

```


QY 391 LPEVTLIVR---SRIVNPKRAVLVTSLNGVAENLTKSMWEPTSTGELIGVQPS 447
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 350 LPELVK--QRNPSSSVYLNRKSKAVFISSSPGYRKSIDVWENTVWGELLVHKPS 407
OY 448 QESTYOOTEKKMHNGKALAEWYLLSTDNL-VTSAMSTFGYVAOGLGKLPWIIRPENRT 506
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 408 YKNYOCTPRNMESKRMAVEIITLLSCDALVTGTGLMSLVEAHGLGKLPMVLNKANLNGT 467
OY 507 TPDPCGRAMSEPCFHSPPFYDCK 531
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 468 AHPHYCVKARSIEPCCSQAATLFHCK 492

RESULT 12
NODZ_RHISN
ID NODZ_RHISN STANDARD: PRT: 322 AA.
AC P5355:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Modulation protein 2 (EC 2.4.1.-).
GN NODZ OR Y4AH.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
Perret X.;
RL "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- FUNCTION: FUCOSYLTRANSFERASE WHICH ADDS THE FUOSE MOIETY OF THE
CC NDO FACTOR ON ITS TERMINAL REDUCING N-ACETYLGLUCOSAMINE END. USES
CC GDP-FUCOSE AS THE DONOR GROUP.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: AE000064; AAB91605.1; -
KW Modulation; Transferase; Glycosyltransferase; Plasmid.
SQ SEQUENCE 322 AA; 36523 MW; 07800CF24CFCFB7B CRC64;

[illegible]

```

QY 437 TGEIGVHPSQEG-YQOTEKMHNC-KALAEYLLSLTDNLV---TSAMSTF 484
      | | | : | | | | | | : : | | : :
Db 241 ---IPKSFRAQDSGPLHSADLGVGGISALVEMYLGLCDYIRFPPTSATFTRY 291

```

RESULT 13

ID	UVRG_HELPJT	STANDARD:	PRT:	594 AA.
AC	092L21:			
DT	30-MAY-2000 (rel. 39, Created)			
DT	30-MAY-2000 (rel. 39, Last sequence update)			
DT	15-JUN-2002 (rel. 41, Last annotation update)			
DE	Exonuclease ABC subunit C.			
GN	UVRG OR JHP0760.			
OS	Helicobacter pylori J99 (Campylobacter pylori J99).			
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group			
CC	Helicobacter.			
OX	NCBI_TaxID=85963;			
NN	111			

RP SEQUENCE FROM N.A

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,

RA Tummino P.J., Caruso A., Iria-Nickelsen M., Mills D M, Ives C
Duch D.K., Noonan B., Gull B.C., deJonge B.L., Carmel G.,

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.
Trust T. J.

"Genomic sequence comparison of two unrelated isolates of the human

RL Nature 397:176-180(1999).

CONCLUSION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS

PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). ATTACHES TO THE
IVRA-IVRB COMPLEX AND THE DAMAGED DNA STRUCTURE

NICKED ON BOTH SIDES OF THE DAMAGED SITE (BY SIMILARITY).

-1- SUBCELLULAR LOCATION: Cytoplasm;c (By similarity).

CC -!- SIMILARITY: CONTAINS 1 UVR DOMAIN.

This SWISS-PROT entry is copyright . It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on

use by non-profit institutions as long as its content is in no way

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC -----
CC -----

DR EMBL; AE001506; AAD06337.1; -.
DR InterPro: IPR001943: IYRH/C

DR InterPro; IPR004791; UVRC.
DR InterPro; IPR001163; UVRC.

DR InterPro; IPR000305; UVR_C_N.

DR Pfam; PF02151; UVR; 1.

```
DR  SMART: SM00465: GTYC: 1
DR  P10D0M; PD0038/0; UVIC_C; 1.
```

```
DR  TIGRFAMS; TIGR00194; uvrc; 1.
DR  PROSITE; PSE0151; 1
DR  PROSITE; PSE0151; 1
```

KW SOS response; Excision nuclease; DNA repair; Complete proteome.

SEQUENCE	AA:	MMW:	OE63C45BA42342E7	CRC64:
594	AA:	68671	MMW:	OE63C45BA42342E7

Query Match 3 28. Score 97. DB 1: 1000000000.

Best Local Similarity 20.8%; Pred. No. 4.9;

100, cups

0 219V LUNOVAL IN INNOBNDL -SGLAS -----GEDEBSCLSKRYQSVHYRK 122

249 AKLYDDLLFAFYGGNNKAVLVKMFMRGKKISSAFEKIHSLNGFDTDAMKQAIINHYS 308

QY 123 PSPYKPSYLISKLRNYEKLHKRCGPTESYKKALKQLDQEHIDGDGECKYVWVISFSGL 182

—


```

Db 309 HLPMPQILLSCSN-----ETLKL-QEFLSHQSKIALSIPKGD 351
QY 183 GNRILSLA-----SVFLYALLTDRVLLVDGRKMDDLFCFCEFLGMSWLLPLDFPMTQFD 237
Db 352 KALIELAMNNAQEIFSQEKTSNEDRLLEARSJLFLNCEVY-----RVLEIFD 399
QY 238 GLNDESRRCYGVW--KNQYIDTEGLSHLYLHLVHDYGDHDKM--FFCEGDQTFICKVP 293
Db 400 TSHRSNSQCVGCVVYENMFQKD---SYRRYHL-KGSNEYDQMSSELLTRALDFAKEPP 455
QY 294 WLIVKTDNYFVPSLMLRFGFDELIN 318
Db 456 -----PMLWIDVGGRAQLN 469

```

RESULT 14

```

MDRB_HUMAN
ID MDRB_HUMAN STANDARD; PRT; 1224 AA.
AC 09BZH6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE WD-repeat protein 11.

```

```

GN WDR11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21427518; PubMed=11536051;
RA Chernova O.B., Hunyadi A., Malaj E., Pan H., Crooks C., Roe B.,
Cowell J.K.;
RT "A novel member of the WD-repeat gene family, WDR11, maps to the 10q26
region and is disrupted by a chromosome translocation in human
glioblastoma cells."
RL Oncogene 20:5378-5392(2001).
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- DISEASE: A form of glioblastoma is characterized by a chromosomal
translocation t(10;19)(q26;q13.3) which involves WDR11 and ZNF320.
CC -1- SIMILARITY: CONTAINS 9 WD REPEATS (TRP-ASP DOMAINS).

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF320223; AAK08064.1; -.
DR MIM; 606417; -.
DR InterPro; IPR001680; WDA0.
DR Pfam; PF00400; WDA0; 4.
DR SMART; SM00320; WDA0; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS50082; WD_REPEATS_2; FALSE_NEG.
DR PROSITE; PS50294; WD_REPEATS_REGION; FALSE_NEG.
KW Repeat; WD repeat; Transmembrane; Proto-oncogene;
KM Chromosomal translocation.
FT REPEAT 59 108 WD 1.
FT REPEAT 111 154 WD 2.
FT REPEAT 354 393 WD 3.
FT REPEAT 471 510 WD 4.
FT REPEAT 566 605 WD 5.
FT REPEAT 708 745 WD 6.
FT REPEAT 747 787 WD 7.
FT REPEAT 793 831 WD 8.
FT REPEAT 893 940 WD 9.
FT TRANSMEM 1127 1147 POTENTIAL.
SQ SEQUENCE 1224 AA; 136683 MW; 9182221ABAEFB4ED CRC64;

```

Query Match

3.2%; Score 97; DB 1; Length 1224;

```

Best Local Similarity 18.9%; Pred. No. 13;
Matches 79; Conservative 61; Mismatches 149; Indels 128; Gaps 18;
QY 152 SYKKALQLODDEHIDGGECKYVWISFSGKNRLLSLASFYLLALLDRVLVDGRKDM 211
Db 824 SMKACFRMDQDELTEPYWCPLYL-----LVPRASIALKAFLIH----- 861
QY 212 DDLFCFPLGMSWL--PLDFPMTDQFDGLNQESSRCYGVWKNQVIDTEGLTSLYHL 269
Db 862 -----QPMNGYSIDISIVDYPENBEIKNLQBDLNSLDIKKLLDPETTLQRCLLV 916
QY 270 VHDYGDHDKMFCEGDQTFICKVPWLIVKTDNYFVPSLMLRFGFDELINLKLPQKATVFH 329
Db 917 SRLYGDSESLHF-----WTVA--HYLHSL--SQEKASSTAPKREA--- 954
QY 330 HLGRYLFRPTNQVGLTRYVEATLSHAD-----EKIGDQ-VAVPEDPRCPGHVNDQI 382
Db 955 -----PRKLSNPDLIDICVDLCENAVYFQKFLQERVNLQEVK-----RSTYDHT 997
QY 383 SSCQKEKILPEVDLVE-----RSRHVNTPKKAVLWLSNAGYAENLKSMEWEPT 435
Db 998 RKCTDQILLGLQTPRAVQLLLETSDADNDHYCDSLKACLVTV-----T 1041
QY 436 STGEIIGVHQPDSQEGYQTEKKM-HNGKALAEWYLLSTDLNLTSA-----WSTFGY 486
Db 1042 SSG-----PSQSTIKLVATNMJANGKLAGYQLLCLIDKAACACRYLQTYGEMNRAM 1094
QY 487 VA-----QGLGGLKPMI--LYPENRTTPDP-----SCGRAMSMEPCFHSPPFYD 529
Db 1095 LAKVLNPEECADVLRRWVDHLCSPQVOKSKALLVLTLGCFPSVAETLHSMRYFD 1151

```

RESULT 15

```

UVR_C_HELPY
ID UVR_C_HELPY STANDARD; PRT; 594 AA.
AC P56428;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exonuclease ABC subunit C.
GN UVR_C OR HP0821.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Loftus B., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA McKenney K., Richardson D., Dodson R., Khaliq H.G., Glodek A.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori."
RL Nature 388:539-547(1997).

```

```

CC -1- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). ATTACHES TO THE
CC UVR-A-UVRB COMPLEX, REPLACING THE MODIFIED BASE(S).
CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVR_A, UVRB AND UVR_C.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE UVR_C FAMILY.
CC -1- SIMILARITY: CONTAINS 1 UVR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compunex Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 10:24:25 ; Search time 4332 Seconds

(without alignments)
1165.481 Million cell updates/sec

Title: US-10-037-311a-2

Perfect score: 1662
Sequence: 1 atgatacgaatcctacag.....ggagacttaagctatgta 1662Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: GenEmbl.*
2: gb_da.*
3: gb_htg.*
4: gb_in.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_un.*
28: em_vi.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rnd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and its derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1632.2	98.2	1768	8	AF154111 Arabidops
2	1401.6	84.3	9318	8	AC005313 Arabidops
3	695.2	41.8	99547	8	AC068197 Genomic s
4	695.2	41.8	112126	8	AC007576 Arabidops
5	642.6	38.7	1698	8	AF223643 Pisum sat
6	607.4	36.5	89301	8	AC006920 Arabidops
7	606	36.5	1512	6	AX412254 Sequence
8	606	36.5	1512	6	AX412452 Sequence
9	606	36.5	1512	6	AX412825 Sequence
10	604	36.3	1602	8	AF417474 Arabidops
11	554.6	33.4	160096	2	AF417475 Arabidops
12	373.6	22.5	147289	2	AP005002 Oryza sat
13	344.4	20.7	134402	8	AC011765 Arabidops
14	344.4	20.7	1706	8	AF417473 Arabidops
15	336.2	20.2	162342	2	AP003458 Oryza sat
16	336.2	20.2	187916	2	AP004687 Oryza sat
17	336.2	20.2	162342	2	AP003458 Oryza sat
18	336	20.2	187916	2	AP004687 Oryza sat
19	229.8	13.8	82771	2	AP003458 Oryza sat
20	229.8	13.8	82771	2	AP003458 Oryza sat
21	150	9.0	147227	8	OSJN00013 Arabidops
22	94	5.7	149120	2	AP005536 Oryza sat
23	93.2	5.6	151133	2	AP004770 Oryza sat
24	93.2	5.6	151133	2	AP004770 Oryza sat
25	93.2	5.6	188228	2	AP004771 Oryza sat
26	71.8	4.3	148669	2	AC091532 Oryza sat
27	55	3.3	150116	2	AP004816 Oryza sat
28	55	3.3	153717	2	AP004841 Oryza sat
29	49	2.9	151163	8	AC079128 Oryza sat
30	48.4	2.9	1141	6	AX083744 Sequence
31	45.4	2.7	144563	9	AC006204 Homo sapi
32	45.4	2.6	7218	6	166494 Sequence 14
33	43.4	2.6	7001	6	AX347354 Sequence
34	43.4	2.6	7001	6	AX349075 Sequence
35	42.4	2.6	176486	8	AP003768 Oryza sat
36	42.2	2.5	186305	9	AL356098 Human DNA
37	41.4	2.5	3625	3	AF081801 Dictyoste
38	41.2	2.5	91001	8	AC013453 Arabidops
39	41.2	2.5	159581	2	AC108667 Homo sapi
40	41.2	2.5	170285	2	AC069237 Homo sapi
41	41	2.5	19395	3	CER0381 Caenorhabdi
42	41	2.5	87081	8	AB073164 Arabidops
43	41	2.5	112684	8	AP002033 Arabidops
44	41	2.5	165458	2	AC009795 Homo sapi
45	41	2.5	171962	9	AC091824 Homo sapi

ALIGNMENTS

RESULT 1
AF154111 1768 bp mRNA linear PLN 28-JUN-1999
LOCUS Arabidopsis thaliana xyloglucan fucosyltransferase (FT1) mRNA,
DEFINITION complete cds.
ACCESSION AF154111
VERSION AF154111.1 GI:5231144
KEYWORDS SOURCE
ORGANISM Arabidopsis thaliana.
Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Perlin,R.M., DeRoche,A.E., Bar-Peled,M., Zeng,W., Norambuena,L.,

Orrellana, A., Raikhel, N.V. and Keegstra, K.
Xyloglucan fucosyltransferase, an enzyme involved in plant cell
wall biosynthesis
JOURNML Science 284 (5422), 1976-1979 (1999) *Jme*
MEDLINE 99301928
PUBMED 10373113
REFERENCE 2 (bases 1 to 1768)
AUTHORS Perrin, R.M., DeRoche, A.E., Bar-Peled, M., Raikhel, N.V. and
Keegstra, K.
TITLE Direct Submission
JOURNML Submitted (25-MAY-1999) MSU-DOE Plant Research Laboratory, Michigan
State University, Room 110 Plant Biology Building, East Lansing, MI
48824, USA
FEATURES
source Location/Qualifiers
1..1768
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="BAC T18E12; C1C11A04"
/clone="EST 191A67"
/clone_lib="lambda ZapII (Dr. J. Kieber, Dr. J. Ecker and
associates)"
/dev_stage="3 day old seedling"
1..1768
/gene="Frl1"
19..1695
/function="transfers Fuc to Gal in alpha-(1,2) linkage on
the plant cell wall polysaccharide xyloglucan"
/codon_start=1
/product="xyloglucan fucosyltransferase"
/protein_id="AAD41092.1"
/db_xref="GI:5231145"
/translation="MDNSYRRSSPIRTTGGSTSVNSELLOKIISSGTMKTRT
FTTCLIVSVIVAFSMIRHQRHSDSNRIRNGFEARVLDAGVPNTNINSKLLJGL
ASGPEDSLSTRYQVHKRPSPYKPSYLLSKLNYEKLHKRCGPTESEYKALKOL
DOEHIDGDECKVWVWISFSGNRIELSAFYLAALLTDRLVDRGKDMDDLECEP
FLGMSMLPLDPMTDQFDGLNDESSRCYGVNQVIDETGLSHLYLPHYGDH
DKMFCEDQPTFIRKVPVLYKTQVFPVSLMLIGPDEDLNKLFEPOKATVFNHNGRY
LEHFNQVWGLTVRYEAVLYSHADEKIGQYRVDEDPGPHQVWDOJSSCTOKEL
FVDTLVRSRHVNTPKHRAVLYTSLNGVAMENLKSMTWETSTIGELTGVHPSOBS
YOOTKKHNCALAEWYLITDMLVTSANSTETVVAQGLGKLPWLTGVPENRRTTP
DPSGGRASMECFSPPTDKAKTGIDTGTVPHVHNCEDISWGLLV"

BASE COUNT 487 a 359 c 378 g 544 t

ORIGIN

Query Match 98.2%; Score 1632.2; DB 8; Length 1768;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1659; Conservative 0; Mismatches 3; Indels 15; Gaps 1;

QY 1 ATGGATCAGATTCTGTACAGAGAAAGATCGTCCGATCAGAACCACTACCGCGGTTCA 60
DB 19 ATGGATCAGAAATTCGTACAGAGAAAGATCGTCCGATCAGAACCACTACCGCGGTTCA 78
QY 61 AAGTCGGTAATTTTCGCAACTACTCTCAATGAGATCTCAGTCCGGATGATGAG 120
DB 79 AAGTCGGTAATTTTCGCAACTACTCTCAATGAGATCTCAGTCCGGATGATGAG 138
QY 121 CTCACGAGAACCTTCACTACTGTTGATAGTCTCTGATAGTATGATGATGATGATG 180
DB 139 CTCACGAGAACCTTCACTACTGTTGATAGTCTCTGATAGTATGATGATGATGATG 198
QY 181 ATCTTTACCAACACCATCTGATTCGAATGGATTATGGTTTCGCGAAGCTAGAGTT 240
DB 199 ATCTTTACCAACACCATCTGATTCGAATGGATTATGGTTTCGCGAAGCTAGAGTT 258
QY 241 CTCGACGCCGAGTTTTCCTCA-----AATTTGATTAAGCTTCTCGGAGGG 285
DB 259 CTCGACGCCGAGTTTTCCTCAAAATGTTTCTAATCAATTCGATTAAGCTTCTCGGAGGG 318
QY 286 CTACTGCTTCTGTTTGTATGAAGATTCCTGCTTAGTAGTACCATCATGTTTCATTAC 345
|||||

DB 319 CTACTGCTTCTGTTTGTATGAAGATTCCTGCTTAGTAGTACCATCATGTTTCATTAC 378
QY 346 CGTAAACCTTCACCTTACAAAGCATCTTCTATCTCATCTTAAGCTTAGAAATACGAA 405
DB 379 GTTAAACCTTCACCTTACAAAGCATCTTCTATCTCATCTTAAGCTTAGAAATACGAA 438
QY 406 AAGCTTCACAAAGCATGTGTCGGGTACTGATCATCTTACAAAGCTTACAAAGCTT 465
DB 439 AAGCTTCACAAAGCATGTGTCGGGTACTGATCATCTTACAAAGCTTACAAAGCTT 498
QY 466 GATCAGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 525
DB 499 GATCAGAACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 558
QY 526 GGTTCAGGAGACGATCTTCTCAGCGTGGTTTTCCTTACGCGCTTTTACCGAT 585
DB 559 GGTTCAGGAGACGATCTTCTCAGCGTGGTTTTCCTTACGCGCTTTTACCGAT 618
QY 586 AGAGCTTGTCTTGTGACCGAGGAAAGACATGATGATGATGATGATGATGATGATG 645
DB 619 AGAGCTTGTCTTGTGACCGAGGAAAGACATGATGATGATGATGATGATGATGATG 678
QY 646 GGTATGCTGTGTTGCTACCTTACATTTCCCTATGACTGATCAGTTGATGATTAAT 705
DB 679 GGTATGCTGTGTTGCTACCTTACATTTCCCTATGACTGATCAGTTGATGATTAAT 738
QY 706 CAAGATCATCTCTGTTTATGATATATGATGATGATGATGATGATGATGATGATGATG 765
DB 739 CAAGATCATCTCTGTTTATGATATATGATGATGATGATGATGATGATGATGATGATG 798
QY 766 ACTTGTCTCATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 825
DB 799 ACTTGTCTCATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 858
QY 826 TTCTGTGAAGAGACCAACATATCATGCGGAATCCCTGGTGGATTTTAAACGAC 885
DB 859 TTCTGTGAAGAGACCAACATATCATGCGGAATCCCTGGTGGATTTTAAACGAC 918
QY 886 AATTACTTGTTCATCTCTGTGTTTAAATACCGGGTTTCATGATGATGATGATGATG 945
DB 919 AATTACTTGTTCATCTCTGTGTTTAAATACCGGGTTTCATGATGATGATGATGATG 978
QY 946 TTCCACAGAAAGGACGCTTTCATCAGTACTAGTATGATGATGATGATGATGATGATG 1005
DB 979 TTCCACAGAAAGGACGCTTTCATCAGTACTAGTATGATGATGATGATGATGATGATG 1038
QY 1006 CAAGTATGCGGCTTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGT 1065
DB 1039 CAAGTATGCGGCTTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGT 1098
QY 1066 ATTGGATTCAGTAAAGATTTTCATGAAAGACCGGGTTCATGATGATGATGATGATG 1125
DB 1099 ATTGGATTCAGTAAAGATTTTCATGAAAGACCGGGTTCATGATGATGATGATGATG 1158
QY 1126 CAGATTCATCTTACTACAAAAGAGAACTCTTCTGAAATAGACACATGATGAG 1185
DB 1159 CAGATTCATCTTACTACAAAAGAGAACTCTTCTGAAATAGACACATGATGAG 1218
QY 1186 AGATCTCGCATGTAAATACCCCAACACAAAGCCGCTGTTTCATGATGATGATGATG 1245
DB 1219 AGATCTCGCATGTAAATACCCCAACACAAAGCCGCTGTTTCATGATGATGATGATG 1278
QY 1246 GGTTCAGGAGAACTTAAAGATATGATGATGATGATGATGATGATGATGATGATGAT 1305
DB 1279 GGTTCAGGAGAACTTAAAGATATGATGATGATGATGATGATGATGATGATGATGAT 1338
QY 1306 ATCGTGTTCATCAGCGGAGCCAAAGAGTTATCAGCAGACCGAAAGAAAGATGAT 1365
DB 1339 ATCGTGTTCATCAGCGGAGCCAAAGAGTTATCAGCAGACCGAAAGAAAGATGAT 1398
QY 1366 GGCAGAGCTTCTGGAATATCTTGTGATGATGATGATGATGATGATGATGATGATG 1425
DB 1399 GGCAGAGCTTCTGGAATATCTTGTGATGATGATGATGATGATGATGATGATGATG 1458
|||||

QY 1426 TGGTACATTGGATGATAGCTCAGGCTCTGGAGGTTTAAAGCCTTGACTCTAT 1485
|||||
DB 1459 TGGTACATTGGATGATAGCTCAGGCTCTGGAGGTTTAAAGCCTTGACTCTAT 1518
|||||
QY 1486 AGACCCGAAAAACCGTACACTCCGATCCTTGTGTGTGGGCTATGTCATGAGCCT 1545
|||||
DB 1519 AGACCCGAAAAACCGTACACTCCGATCCTTGTGTGTGGGCTATGTCATGAGCCT 1578
|||||
QY 1546 TGGTCCACTGCGCTCCATTCATGATGTAAGGAAAAAGCGGATTTGACACGGGACA 1605
|||||
DB 1579 TGGTCCACTGCGCTCCATTCATGATGTAAGGAAAAAGCGGATTTGACACGGGACA 1638
|||||
QY 1606 CTAGTTCCTCATGTGAGCATTTGTGAGCATATCAGCTGGGACCTTAAGCTAGTATGA 1662
|||||
DB 1639 CTAGTTCCTCATGTGAGCATTTGTGAGCATATCAGCTGGGACCTTAAGCTAGTATGA 1695
|||||
RESULT 2
AC005313 91318 bp DNA linear PLN 11-MAR-2002
LOCUS
DEFINITION Arabidopsis thaliana chromosome 2 clone T18E12 map C1C11A04,
complete sequence.
AC005313
VERSION AC005313.3 GI:20197340
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 91318)
Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M.,
Shen,M., Ronging,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
Unpublished
2 (bases 1 to 91318)
Lin,X.
DIRECT SUBMISSION
REFERENCE
AUTHORS Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
TITLE Medical Center Dr., Rockville, MD 20850, USA
JOURNAL 3 (bases 1 to 91318)
TOWN/C.D. and Kaul,S.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr., Rockville, MD 20850, USA, cdtonet@igr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6558460.
FEATURES
COMMENT Location/Qualifiers
FEATURES
SOURCE 1. 91318
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C1C11A04"
/clone="T18E12"
repeat_region complement(2646..2678)
/rpl_family="(A)n"
repeat_region complement(2800..2854)
/rpl_family="(AT-rich)"
repeat_region complement(3183..3217)
/rpl_family="(AT-rich)"
repeat_region complement(5635..6198)
/note="T18E12.5; predicted by genscan"
gene /gene="At2g03320"
complement(join(<5635..5876,6135..>6198))
/note="T18E12.5; predicted by genscan"
mrna /gene="At2g03320"
complement(join(5635..5876,6135..6198))
/gene="At2g03320"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA15037.1"
/db_xref="GI:20197347"
CDS /translation="MLINLIKINITYHYVFLCMAKVSSEKLSFRKAAREPNLPPIAA
ATTLISPOGRVDFVMDNEAOEIRNVSPISOELPSSSTISHSNERSNLSRSIVIO"

repeat_region complement(6171..6195)
/rpl_family="(AT-rich)"
repeat_region complement(6248..6271)
/rpl_family="(TA)n"
repeat_region complement(6305..6371)
/rpl_family="(AT-rich)"
gene /gene="At2g03310"
/note="T18E12.6; predicted by genscan"
mrna <6491..>6549
/gene="At2g03310"
6491..6549
/gene="At2g03310"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA15038.1"
/db_xref="GI:20197348"
/translation="MNRAEAQEEFSEWVVIQTSKSPSPRTDASPNSPILOPSHR
RENPTIINDDSDSVSPSTVNLMLPWRVLETKKRLIKNSVCIFQAVERRVCSYLRKR
VFWSLTIIGFSLVLSLVYKLVWWRRLHEKLRPILLLLREKDYCVN"
8920..8985
/rpl_family="(AT-rich)"
repeat_region 9006..9072
/rpl_family="(AT-rich)"
repeat_region 9167..9187
/rpl_family="(AT-rich)"
gene 9358..10112
/gene="At2g03300"
/note="T18E12.2; domain signature TIR exists, suggestive
of a disease resistance protein"
mrna join(<9358..9563,9631..9876,9953..>10112)
/gene="At2g03300"
join(9358..9563,9631..9876,9953..10112)
/gene="At2g03300"
/codon_start=1
/product="disease resistance protein (TIR class),
putative"
/protein_id="AA15034.1"
/db_xref="GI:20197344"
/translation="MTFSPPOVFLNWRGEOLRSPVSHLIDAEFNEINEFVUKYEO
RGKDLNLFRIQESKIALALESTRYTESWCMDLVKIKKLADKRLHVPYKVK
VEDVRKQTEGDNFWTLAKVSDQIKKMEALECIPNKGSLIGDKSDEADPIKEV
VKAVOCVATATGLEEENHFGKKRKDCCKLPDLKRSRTKTL"
repeat_region complement(11109..11139)
/rpl_family="(AT-rich)"
repeat_region complement(11148..11178)
/rpl_family="(TA)n"
repeat_region complement(11205..11293)
/rpl_family="(AT-rich)"
gene complement(113046..14158)
/gene="At2g03290"
/note="T18E12.3"
mrna 13 complement(join(<13046..13290,13406..13482,13662..13871,
13949..>14158))
/gene="At2g03290"
complement(join(13146..13290,13406..13482,13662..13871,
13949..14158))
/gene="At2g03290"
/codon_start=1
/product="putative Golgi-associated membrane trafficking
protein"
/protein_id="AA15035.1"
/db_xref="GI:20197345"
/translation="MDCRSSILLLIALLSPRTLMSRYELKSKTYCIGCEETHENAM
SICKYFFIVNPNEDHPLPDSHKITIVKVPDQKNLHBDKVEAGQFSFTAYENGSSVYA
CITAIIDYKPEPTTLIDFDWKGTGHSKEMTNVAKKSOYDMMEYQVTLMDTVISIEEM
YYLREDEEMOELNRSTNSKMAWSPFSLVVCISVAGIQFWHLKTFPEKKKLI"
repeat_region complement(14743..14793)
/rpl_family="(AT-rich)"
gene 15341..17331
/gene="At2g03280"
/note="T18E12.4"
mrna join(<15341..15436,15491..15590,15680..15753,15817..15905,

```
15987..16149,16242..16531,16607..16811,16886..17043,
17109..>17331)
/gene="At2g03280"
join(15341..15436,15491..15590,15680..15753,15817..15905,
15987..16149,16242..16531,16607..16811,16886..17043,
17109..17331)
/gene="At2g03280"
/codon.start=1
/product="similar to axi 1 protein from Nicotiana tabacum"
/protein_id="AA15036.1"
/db_xref="GI:20197346"
/translation="MSVVGDRESESTLMRSDYKAPPSQAIPKARLQLFMHSOITFGLT
NHISFSLPVQSVLPPLPRNYSNGLIVSCGGLNQMAARVARDIMQTVARLN
LTVPELDKSSFMADISDFEDIDIKHIDSLRDEYVIRLPRKYSKXGKFLFEP
PVSNDKXYLQOVLPRFSKRYIHFRSPSLANNGLSIDLRNRVAFGLRFT
PRIEGLSKIVRILQOGRSFVALHRYMDMLASGCHGTGCTDEAEELKMRVAPW
MREKIVEERVRVQGLCPITPBEAVLVKALGFQDDIYIAGETFGGARLALKE
SPRLVKEMLDLPTELQFQNHSSQMALDFIVSAVNTFIPYYGMMAVYEGHR
YLGFRKTLIDRRKRLVELLDLHNKNTLSWDOFAVAVKADHOGRRGEPTHRKVISVR
KEEDFYANPOECIS"
complement(17450..19797)
/gene="At2g03270"
/notes="11812.1: supported by cDNA:
gi_15027926_gb_A1045820_1_"
complement(join(<17450..19527,19705..19797))
/gene="At2g03270"
complement(17590..19509)
/gene="At2g03270"
/codon.start=1
/product="putative helicase"
/protein_id="AA15035.1"
/db_xref="GI:20197343"
/translation="MARKMSLEAFVSTAPLIDMEKRAEISMSLTSGASRIETAOAK
GTTILNKCVQVQGLMGRSLIEFQSNQGVLPARKFGNHVVLKLNKSPDGLA
OGVYRLKDSISITVPEDEVPEEGNTSLREKLEKANEYRMRKDTLQLSKGTGGA
SDVPLVEGERPQSKDYSPENKKNIDQSKQALITRALSQVDFLHGPEGTGK
TTTYVELQEVKRGSKILACASNIANDIVEKLVPHKXKLVNKHVHARLPVQDS
ALDAQVKNADVILTLTGALTRKLDNRFDLVIIDEQAQALEVCWIALKSGRSCIL
GDHLQPTTIOASAEKRGIGRTLEFRLADYGEIKSMILVOYRMHLEINMSKEL
YDNKLTASHSVASHMFLDEWTKSSSEATLLVDPAGCMEMKKEESTVNEGSA
EVAMAHAKRLMESVOPSDIGILTPYAAVWLLRIILGKEKELKDMISYDQOGE
KEATITSMVRNSKKEGVGLKDDRRMVAVTRSRQCCIVCDITETVSSDAFLKMIET
FEHGEYLSASEYN"
21248..25474
/gene="At2g03260"
/notes="11812.7"
join(<21248..21521,21643..21984,22920..23311,23552..23681,
23760..23880,23973..24231,24320..24424,24497..24691,
24770..25066,25150..25286,25375..>25474)
/gene="At2g03260"
join(21248..21521,21643..21984,22920..23311,23552..23681,
23760..23880,23973..24231,24320..24424,24497..24691,
24770..25066,25150..25286,25375..25474)
/gene="At2g03260"
/codon.start=1
/product="unknown protein"
/protein_id="AA15032.1"
/db_xref="GI:20197341"
/translation="MKRGKELSSOMVOEQOAVVNDYLTKLLEIKLKKEKNPPPP
PHAAVEGEGISRKMTLYRARSGLYOPRGKROSSROSSSEIDIEGKAPITYKST
HGLPTPLMAERGEVELYFPRRLDEFNRKREKYEKWDKIMLNKQMDALT
AFRKVENPVGMWEKTEVMTSLADIAVSTAIAASTARTKMPRAQAMEAIO
EGSFRENEDHDGSRGATGVKTSLSLNTMRAPAPIVLDHIKLNNTKAPRSTI
KGVNLSSONEILFNRLONLEVEKELFAVEYEROKRLKSYSLVLAFLSKILKY
DKLMLKIORVESTIKHFANGHRKGNILRPMOKRKHVSTSTGSAACISLIVA
LVAILTRKTMPERENHTYMNTEPLYSGLGFIHLTHTWKVIDIYWKRYRWYARF
GCGQGTGLGYROYLFLGTTIGTRALLCVGLNDLMEVNPKRKNRPLLELLPLIALV
FVYLIMFHLIYSTRFTLTCLDHCIAAPLVITLPDFLQDULTISQVALSINFY
ICYIGWDFKRNQTCASEIYISLYIVASLYLSLQDCMRMIEERSLDGYNV
KLYLTVAISLRTAGYEVNKNPTSHLVLVLAGSSIIAAVCTWDFHMDGLNK
TSKRNRLRDLILPOKRYVFIAMLVNVLFAFMIQTLNFEFLKQDTLAVVASLE
IMRGWMPFRVENEHLNNYKFRAPKSVPLPENVYEDDKD"
```

```
repeat_region complement(26234..26278)
Query Match 84.3%, Score 1401.6; DB 8; Length 91318;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 255 TTTCCCAATTCGTGATGATGCTTTCGGAGGCGTACTGCTTCGTGGTTTGTGTAAGATTC 314
Db 41772 TTTTTCACATTCGTGATGATGCTTTCGGAGGCGTACTGCTTCGTGGTTTGTGTAAGATTC 41831

QY 315 TTGCCTTAGTAGTACCAATCAGTTCATTTACCGTAACCTTCACCTTACAAAGCCATCTTC 374
Db 41832 TTGCCTTAGTAGTACCAATCAGTTCATTTACCGTAACCTTCACCTTACAAAGCCATCTTC 41891

QY 375 TTATTCATCTCTAACTTAAACATACGAAAGCTTCACAGCGATGTGCTCCGGGTAC 434
Db 41892 TTATTCATCTCTAACTTAAACATACGAAAGCTTCACAGCGATGTGCTCCGGGTAC 41951

QY 435 TGAATCTTACAGAAAGCTTAAACAACTGATCAAGAACATATGATGATGATGATG 494
Db 41952 TGAATCTTACAGAAAGCTTAAACAACTGATCAAGAACATATGATGATGATGATG 42011

QY 495 ATGCAATATGTGTGTGATTTCTTTTAAAGCGCTTAGGAAACAGATACCTTCTACG 554
Db 42012 ATGCAATATGTGTGTGATTTCTTTTAAAGCGCTTAGGAAACAGATACCTTCTACG 42071

QY 555 CTCGGTTTCTTAAAGCGCTTTTAAAGCGATAGAGTCTGTTTTCACCGAGGAAAGA 614
Db 42072 CTCGGTTTCTTAAAGCGCTTTTAAAGCGATAGAGTCTGTTTTCACCGAGGAAAGA 42131

QY 615 CATGAGATGATCTTTTTCGAGCGCTTTCGATGATGATGATGATGATGATGATGATG 674
Db 42132 CATGAGATGATCTTTTTCGAGCGCTTTCGATGATGATGATGATGATGATGATGATG 42191

QY 675 CCTCATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 734
Db 42192 CCTCATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 42251

QY 735 GGTAAAGATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 794
Db 42252 GGTAAAGATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42311

QY 795 TCATGATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 854
Db 42312 TCATGATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42371

QY 855 GAAAGTCCCTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 914
Db 42372 GAAAGTCCCTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42431

QY 915 ACCGGTTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 974
Db 42432 ACCGGTTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 42491

QY 975 CTTAGTAGTATCTTTTACCCCACTAACCAAGTATGGGCTTAGTACTAGATAC 1034
Db 42492 CTTAGTAGTATCTTTTACCCCACTAACCAAGTATGGGCTTAGTACTAGATAC 42551

QY 1035 CGAAGCTTACCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1094
Db 42552 CGAAGCTTACCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42611

QY 1095 AGACCCGGGTCATTTACGATGATGATGATGATGATGATGATGATGATGATGATGATG 1154
Db 42612 AGACCCGGGTCATTTACGATGATGATGATGATGATGATGATGATGATGATGATGATG 42671

QY 1155 ACTTCTACCTGAAGTAGACACACTAGTGAAGAGATCTGCCATGTTAATACCCCAAGA 1214
Db 42672 ACTTCTACCTGAAGTAGACACACTAGTGAAGAGATCTGCCATGTTAATACCCCAAGA 42731

QY 1215 CAAAGCGGTCTTGTACATCTTTGAACGGGGTTCACGGGAGAACTTAAAGAGTATG 1274
Db 42732 CAAAGCGGTCTTGTACATCTTTGAACGGGGTTCACGGGAGAACTTAAAGAGTATG 42791
```

OY	1275	TTGGGAATATCCGACATCAACTGAGGAATATATCGGCTGTCTATCGACCGCAAGAAAG	1334
Db	42792	TTGGGAATATCCGACATCAACGAGGAATATATCGGCTGTCTATCGACCGCAAGAAAG	42851
OY	1335	TTATAGCAGACCCGAAAAAAGATCATTAATGGCAAGCTCTTGGGGAATATATCTTTT	1394
Db	42852	TTATAGCAGACCCGAAAAAAGATCATTAATGGCAAGCTCTTGGGGAATATATCTTTT	42911
OY	1395	GAGTTTGACAGATAATCTTGTGACAAAGTCTTGCTACATTTGGATATGTAGCTCAAG	1454
Db	42912	GAGTTTGACAGATAATCTTGTGACAAAGTCTTGCTACATTTGGATATGTAGCTCAAG	42971
OY	1455	TCTTGGAGGTTTAAAGCCTTGATATCTATATGACCCGAAACGCTACAACTCCGATCC	1514
Db	42972	TCTTGGAGGTTTAAAGCCTTGATATCTATATGACCCGAAACGCTACAACTCCGATCC	43031
OY	1515	TTTCGTGGTGCGGCTATATGTCATGAGAGCCGTGTTACATCGCCCTCATCTATGATTG	1574
Db	43032	TTTCGTGGTGCGGCTATATGTCATGAGAGCCGTGTTACATCGCCCTCATCTATGATTG	43091
OY	1575	TAAAGCGAAAAAGGCTATTGACACCGGAACACTAGTTCCTCATGTGAGACATTTGAGGA	1634
Db	43092	TAAAGCGAAAAAGGCTATTGACACCGGAACACTAGTTCCTCATGTGAGACATTTGAGGA	43151
OY	1635	TATCAGCTGGGAGCTTAAAGCTATGATGA	1662
Db	43152	TATCAGCTGGGAGCTTAAAGCTATGATGA	43179

TITLE	Yu G., Davis R., Federpiel N., Theologis A. and Ecker J.
JURNAL	Submitted (28-7/09-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
ENTRY	On May 19, 2000 this sequence version replaced gi:7767679.
SOURCE	location/Qualifiers
	1..99547
	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"
	/chromosome="1"
	/clone="F16A14"
	join(40..972,1061..1201,1297..1578)
	/note="unknown protein; similar to EST gb AI97084.1"
	/codon_start=1
	/evidence="not_experimental"
	/product="F16A14.1"
	/protein_id="AAFW9391.1"
	/db_xref="gi:8778383"
	/translation="MPGFDRISELPESLISQILLHPKASVKTSTSRKNMLNVN PEGLDNCDFEPÖNNNEKLIDIFDRFLQFNENSLRQFKYVDYSRDKI KFSDSIGGA ISRGIRLVDEPENTYYRADDCIDIPCEFPFMNLNYSGKITLSLKLSYSGEDGFVYV LPCLTFEMILREVWMSSTNMLEKLVSQCPAYEEIYLHDKDLVTRRSLSKRFSI PRHLTSLFRRYTOFEETIDAPGLEWMSLKADHPRIYVKNTLSLPMTDIDIKFLVGGS VMEDPEDLPKREIREDPLTGTSVVRHMYISMNTYKALDIYSKVELIKRFNNLSVEEAAA PFSSLLQCPALFLESFPMUKLLILETCQPVMEFFELVNPFCFVSTLEIVETLKGFL DWGEQMDIASYLENSAVLKKRLILTSPMGCPQHYSSESDIYEELMKLTRKPSRCQIIIR C"
CDS	join(2702..3920,4012..4193,4279..4347,4450..4743,
CDS	

RESULT 3	
AC068197	
LOCUS	
DEFINITION	AC068197 99547 bp DNA linear PLN 28-JUN-2000
ACCESSION	Genomic sequence for Arabidopsis thaliana BAC F16M14 from
VERSION	AC068197
KEYWORDS	Chromosome 1, complete sequence.
SOURCE	AC068197.4 GI:7940274
ORGANISM	HTG.
	Arabidopsis thaliana.
	Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 99547)

Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hosson, C.,

REFERENCE

AUTHORS

Khan, S., Kim, C., Altafi, H., Babi, O., Chin, C., Chiu, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Hwang, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Federspiel, N. A., Theologis, A. and Eckert, J. R.
Genomic sequence for Arabidopsis thaliana BAC FL614 from

chromosome I
Unpublished
2 (bases 1 to 99547)
Ecker J.R.
Direct Submission
Submitted (30-APR-2000) Arabidopsis thaliana Genome Center.

Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 99547)

AUTHORS Ecker, J.R.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 36th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 99547)
AUTHORS Cheuk, R., Shinn, P., Brooks, S., Buehler, E., Chao, O.,

Johnson-Hopon, C., Khan, S., Kim, C., Altafi, H., Bel, B., Chin, C., Chiu, J., Choi, E., Conn, L., Conmay, A., Gonzalez, A., Hansen, E., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thavari, A., Tortiumi, M., Vaysberg, M.

ENTRY	source	source	source
YU.G.	Davis, R., Federici, N., Theologis, A. and Ecker, J.		
Direct Submission			
Submitted (28-JUN-2000)	Arabidopsis thaliana Genome Center,		
Department of Biology, University of Pennsylvania, 38th and			
Hamilton Walk, Philadelphia, PA 19104-6018, USA			
On May 19, 2000 this sequence version replaced gi:7767679.			
Location/Qualifiers			
1.99547			
/organism="Arabidopsis thaliana"			
/db_xref="taxon:3702"			
/chromosome="1"			
/clone="F16A14"			
join(40..972,1061..1201,1297..1578)			
/note="unknown protein; similar to Est gb A197084.1"			
/codon_start=1			
/evidence=not_experimental			
/product="F16A14.1"			
/protein_id="AA079391.1"			
/db_xref="gi:8778383"			
/translation="MAGFDRISELPELSIQILHLPTKASVTSVLTSPKRLNMLNVL			
PGLINDGDFPQNNNEKLLIDFDLEDFQNNESRLPKYDSEKDIKTSKIDGAPN			
ISGRIVLDESNTYTRADDCIDPCLIEPMLNISCKTIVSLKLSGLSDGEPVY			
LPCLFMHLREVMYDSSGTMNLEKYSVCPVLEELIYLDKLVATVRSSLSKRFSD			
PERHLSLFRVYQTFEIDAGLEKYSMLKADHPRIYVNLTLSPMLDIDIKFLVGGF			
FMFDEDELPKRNELIDFLTGISVSHVIMSHNTYKALDLYSKGLIKFNLSVEEAA			
PFMSJEDLPALPESEFPMLKHLILETECPVEMKEPFLVNPQCFVSLHEVELKGF			
DMGEBDMKIASFLENSAVLKLLSLFMGCGOHSESDIYELMLKLRSPRCQIIIR			
C"			
Join(2702..3920,4012..4193,4279..4347,4450..4743,			
4956..5165,5257..5493)			
/note="hypothetical protein"			
/codon_start=1			
/evidence=not_experimental			
/product="F16A14.2"			
/protein_id="AA079392.1"			
/db_xref="gi:8778384"			
/translation="MYSRRLEDELERYRYSEMKGDTGRKVKITSESLFRPCFYIDRKRD			
YQFDLLHNASGIGSSSRKTDGDKARILALERYMKYLRGSRGSPSPDVSILPK			
EEFTKMKMSTLSTTEGEFETITENSSPHYIKAEKPPVIGDSSRGSEERLKFSDKDP			
PEFSEEDKSYPARKPCPLVSGAKGDEDPQIRLISGHSFATYPOKILVYISGANGD			
YVHPKGLILANKKTPFNKTKRYKAGESSKTRBDLIKGFPHKVTYPLMNRSLCFPTGF			
ALVDFKMEGFRNATYMFDKHFEVYSGQKRHDHILTRDGDLYGMVAQDDYVSRTH			
GDHLKRQDLSVSGKEADEQKTFETVLSNLELVTYSNLMQMESIYKQTSVLSLR			
RAKNEEDININHNKMSIMQOTARDYLASIYEEHKASQHLAEQKREYDENVLDKCC			
QAKNETERRKILQOMKHNLMATQEQNKADEEMRLAEQOQREDELKQVRELSEKID			
AEQALEETERRRGDLYGMHMOEGEGEDSKITKIEKTKELKLEKEDMDYOSLVO			
TLVYHGTNLEODARKALIRSMELTTRVATYVIRGMGALDELPKFKVAKERYPAVE			
ADKAEELICSLMEELHSGSANHPKYVKESGITAEBELNEDEKLOELKELGEEVYAA			
VTQALKENETVSGSRITVPELWMFKQRKASIEGVYVIVSNKQKKPKPRR"			
complete(6320..8971)			
/note="hypothetical protein"			
/codon_start=1			
/evidence=not_experimental			
/product="F16A14.3"			
/protein_id="AA079418.1"			
/db_xref="gi:8778410"			
/translation="MRVPPIILSHVRGLIRQGGSSREYVVPALARTNLITISHQVAK			
EGTPEPGLALELIDGLVRLVSNMKDPPYLALSPKRLIEGNTVLPSSOAVAVIITCG			
AGLDKDLTPELIVRGDEGRGFSVMDLKAIGMEQSLVLLRVSTALKAAYANLD			
MEDELIDLEPRAYVSLGRAPDITKALNPLISMIVISGRDMVYGFPELIERGLADAAH			
TYVILVQALMRDDKELEKRLSLISFTRNPPCYFINPLEGICLNMQDIAFLFILO			
PLRDLNLSIDSLGIARVYRGGLCTEMRLIEDASVVLDEKHGIDPDVYVSAIIIE			
GIDRKMVNPVPAVFNKMLKRRKRLNCTVYSSIIQCYQOMNSEADLPERPETNIN			
SLDRVCYVPAFDALGKLEVEALELEFEMQKGIADPVINNTLIGCCLOGCSDSPA			
EDLMLEMGQGTGPDIVIVYVNLAGLVAQAEFELTKMENRGVAPTYVTHNVIY			
EGLDAGELGDAEAPVLESLHKSRSNDSVNGKCAAGCLDHAEREFTLEFPLPKSV			
YVLTFTSLCAEDYISKODLIDRMKIKYKPEKSMQKILGACRCNNRKAKPEPE			
ILVTKTIPDLPTFTTITNTYCRNLNPKQATALEBDKRRQVADVYVSYLNSDPE			
LDMKKEMAFADYIDPVVYVYTIMNKYGLNDLKVVALFQMKRRREIVPDVYTVLL			
NKRPRLNISREKAKADKAPDVYVTVLLIDMCKTGIDIGEARIPDOMIESGVDPADAP			
YVALIACCCKMGYLEAKMIFDRLMIESGVKDPVYVYALLAGCCRNFGVFLKAVLVE			
MLEKIRIPKRLSIVHYAKLKAGLAR"			
Join(9332..9383,9480..10419,10934..11089,11318..11412-			

QY 1096 GACCCGGTCCATTTACACATGTGATGATGATTCATCTTGTACTCAAAAAGAGAA 1155
 DB 94816 CAACTAAATATTTCCACAGCATGTGACCAATCGTAGCATGACAAAGAGAA 94875
 QY 1156 CTTTACTGAAATGACACACTAGTG-----AGAACTCTCCGCAATGTTATACCCC 1209
 DB 94876 CTGTATCTGAATTTGCTGCAAGAAAGACACAGTAATATCAATCAATCCCTCC 94935
 QY 1210 AACACAAAGCGTCTGTGACATCTTTGAACGGGGTTACGGGAGAACTTAAGAT 1269
 DB 94936 AAACCTAAAGCTGTTCTGTGATCATCTGTAATCCAGTAATCTCAATTAAGAG 94995
 QY 1270 ATGATTTGGGAATATCCGACATCAACTGAGAAATCATCGTGTTCATCAGCCGAGCCAA 1329
 DB 94996 ATGATTTGGGAATATCCGACATCAACTGAGAAATCATCGTGTTCATCAGCCGAGCCAA 1329
 QY 1330 CAAGCTTTCAGCAGACCCGAAAAAAGATGATTAATGCCAAAGCTCTTGGCGAATGAT 1389
 DB 95056 GAAAGCTTTCAGCAGACCCGAAAAAAGCTTTCAGCAGACCCGCGAGAGATGAT 95115
 QY 1390 CTTTGTGATTTGACAGATATCTTGTGACAAAGTGGTGTGATCTGATGATATGATGCT 1449
 DB 95116 CTTTGTGATTTGACAGATATCTTGTGACAAAGTGGTGTGATCTGATGATATGATGCT 95175
 QY 1450 CAAGCTTTCAGCAGACCCGAAAAAAGCTTTCAGCAGACCCGCGAGAGATGAT 1509
 DB 95176 CAAGCTTTCAGCAGACCCGAAAAAAGCTTTCAGCAGACCCGCGAGAGATGAT 95235
 QY 1510 GATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1569
 DB 95236 GATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 95295
 QY 1570 GATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1629
 DB 95296 GATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 95355
 QY 1630 GAGATATATCAG--CTGGGAGCTTAAAGCTAGT 1658
 DB 95356 GAGATATATCAG--CTGGGAGCTTAAAGCTAGT 95387
 RESULT 4
 AC007576 112126 bp DNA linear PLN 16-JUN-1999
 LOCUS Arabidopsis thaliana chromosome I BAC F7A19 genomic sequence,
 DEFINITION complete sequence.
 ACCESSION AC007576
 VERSION AC007576.3 GI:5019264
 KEYWORDS
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 112126)
 AUTHORS Federiciel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
 Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
 Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
 Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vytotskaia,V.,
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 JOURNAL Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 TITLE Direct Submission
 JOURNAL Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 REFERENCE 2 (bases 1 to 112126)
 AUTHORS Federiciel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
 Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
 Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
 Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vytotskaia,V.,
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 JOURNAL Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 TITLE Direct Submission
 JOURNAL Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 REFERENCE 3 (bases 1 to 112126)

AUTHORS
 Federiciel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
 Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
 Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
 Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vytotskaia,V.,
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 JOURNAL Submitted (08-JUN-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 REFERENCE 4 (bases 1 to 112126)
 AUTHORS Federiciel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
 Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
 Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
 Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vytotskaia,V.,
 Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
 JOURNAL Submitted (16-JUN-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 COMMENT On Jun 8, 1999 this sequence version replaced gi:4887737.
 e-mail for correspondence: arabsequence.stanford.edu
 Genes with similarity to proteins in the databases are described as
 'putative', '-like' or 'similar to'. Genes that have EST
 similarity but no significant protein similarity are described as
 'unknown proteins'. Genes that are annotated based only on gene
 prediction software are described as 'hypothetical proteins'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
 http://genomic.stanford.edu/~chris/GENSCANW.html), Fexa (V.Solovayev
 & A.Salamov, Sanger Centre, http://genomic.sanger.ac.uk/),
 NetPlantGene (S.M. Hebsgaard, et al., CMS, Technical University of
 Denmark, http://www.cbs.dtu.dk/NetPlantGene.html) and
 eMotif (Nevill-Manning, C.G. Wu, T.D. & Brutlag, D.L.,
 http://motif.stanford.edu/projects.html).
 FEATURES
 source
 1..112126
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="1"
 /map="q2358"
 /clone="F7A19"
 /cdate="2806"
 /gene="F7A19.1"
 join(1240..1265,1351..1448,1540..1697,1788..1887,
 1976..2334,2426..2507,2612..2806)
 /gene="F7A19.1"
 /note="Hypothetical protein"
 /codon_start=1
 /protein_id="AAD39278.1"
 /db_xref="GI:5080768"
 /db_xref="GI:5080769"
 /translating="MDTLIKOTRRKHPASQKIREVGSSTREKRVASRFEDKK
 KPSMWLOQFQSRWGSQSYDPIDGMDHAAVAAYATPEETWLEHYHGLEIGPS
 SSRSKRSRELLPLEEPRSLRRPQSLFSDFTKDKPTLRSPMKSSSVKTF
 MNLKGDKHKKONQDESEKHEKROKRVSEPRLOPLRTPESEPRAPPPPLPSPLR
 LPPEETKROSSEHRSKDDSTADMEKAEKLSIKARYKLRKIDLMKAKREKARRK
 LTISEQSLDEQRKRRKGLQRFREDTEITLQIAGARAQKDEKRSREKVEKAGVTRS
 TGKLPGNACPF"
 4106..4769
 /gene="F7A19.2"
 4106..4769
 /gene="F7A19.2"
 4143..4610
 /gene="F7A19.2"
 /note="Unknown protein; location of ESTs 128A1477,
 gb|1944625 and 128A14XP 3', gb|A1059841"
 /codon_start=1
 /protein_id="AAD39279.1"
 /db_xref="GI:5080769"
 /translating="MNFISDOVKKLSSTPEEPDHNKPEVETETATRPATNAELMASA
 KVAEAAQAAARNESDKLDKGVAGASADIIIDAAEKYQKPEKSSGQYLDKAKRYLN
 DYESSHSTGAGCPPPTPSQAEPAASQPEPAKAKKDESGGIGIGYAKKAQGFLLK"

```
gene      complement(5755..8975)
CDS       /gene="F7A19.3"
          /complement(join(5755..7820,7930..8641,8781..8975))
          /gene="F7A19.3"
          /note="Hypothetical protein"
          /codon_start=1
          /db_xref="GI:5080770"
          /translation="MGTKYCHESIGCYHHSMGDLNKESSNCGRPLFYGDNKTASN
NOCYNNNGFTSGTTEGDKDYELHRYRTQKDMDELKRFQFNKEMQVLAISUSQ
AIRNDYRKAKIRSFPLANSVYIDPMSVVEDKSNQSGVSNQWNGASKEVE
VSEVRPTKIRKMTDLCLPADEYIDNEVEVLEKHRYCSTSQLPNGARKESRIDG
VRIGYSSRSNLADLNEPDAOEANEFAYGSRDLRNGEFOGHIDYKSTNSGVR
EHIPIVLPQDLNGKPKVYMSHOPRLTDHNGTHKSAAPEOPAKPLDISJQMOML
SSORVMSLPNGSPGPKAVYVHERFTIDLEADDTTNSHEAVTAKPLDLSLSHQAP
LYPIYRFAVAMNHLSWQNSRSEFPKQVSAORYPLANSDDLIGNAQKQGLD
RKQFESNRFDSGCCNSRLNNMTNFCSTSKSVAGTGYNPNPGSGSDYSSEKF
VLDLNLNTLNTSVVEVRKDEHLATLFLMKPKSVCSNSELADKWSLKSDAVPSS
SKPLDIRKADKQNTIMLERLKSGCSNPNVEKIDANKEIPLAYKADQSNVKG
KHYVYMLDMNEPCDPLMDEDOOTEQETRVASNRQIDENILIGDEDEGNCVPA
SRLNSKAPMIDLETVPESDEEDGNTSGEKRSFYKSTIEETLEKPPPEKLAETI
VAIYSACIDREVEVAVSSVSFTIILHMAPYVNTKHEMLDKIDTFESNORSIEDI
DYESWTIOLPDISQDYHPRKPLVBNLKEETTGATLTALYQRPGRNARKKQKRD
OPDILPGLSLSKHEVTEDIOMFDGFMRTGSRWTPGLTRKTKTSRGRPRRAITIPA
EPVYCPVPAPEPPPSVOQHVSQNSNGETEMCLDRSFAGMKTRPRRRCPSAS
TTTATTSNORSSSHAPFT"
16683..17720
/gene="F7A19.4"
/note="EIF-5A"
CDS       /join(16683..16802,17074..17196,17290..17346,17420..17494,
17619..17720)
/gene="F7A19.4"
/product="Initiation factor 5A-4"
/protein_id="AAD39281.1"
/db_xref="GI:5080772"
/translation="MSDEHHPESSDGASKTYPQOAGTIRKNGYIVIKRRCRVVEV
STRTKGHGHAACHFVAIDIFTSKRLEDIVPSHNCADVPVHNRDYLQIDISEDYS
LTDNGSTKDKDLKPNDDTLQOIKSGPDDGDLVVSAMGEQINALDKP"
19674..22095
/gene="F7A19.5"
/join(19674..20007,20763..21468,21551..21715,21837..22095)
/gene="F7A19.5"
/note="Similar to DNA-binding proteins"
/codon_start=1
/protein_id="AAD39282.1"
/db_xref="GI:5080772"
/translation="MSEKEEAPSTKSGAPSRPLTSLPPRPFSEMPFNGVGSPPG
MTIYSNMPDSDDEFPSFSQLLAGAMSSPATIAAAAAATASDYORLGGTSSSGDVD
PFRKONRTAYVLDICNTVQGSYGHTHQOLAQVTAQVQVQANANMOPOTEYPPSOVO
SFSSGOAQIPISAPLPAQRETSVDTIIEHRSQOPLNDKRPADGYNMKTQOKOVKS
EFPRIYKCTGCTGCPVKKVERSLDQVTEIITYGOHHEPDPONTKRNKNPTNINIG
SINNNGSSSELGASQFOFNTSSNKTQRHOHVASQATTEHLSSESDGEEKNPTING
REKDENEPDKRSTRTEVRISEPAASHRVTEPRITVOTJSEVDLLDDGKRMKRYGQ
KVVGNPYPRIYCTTGGCGVRKHVEAATDPAAYVTYTGKHNHDLPAKSSSHAA
AAAOIRPNRPGGLANLMOOQOQOPVAKRLKEQTT"
complement(22794..24934)
/gene="F7A19.6"
complement(join(22794..22904,23002..23124,23218..23412,
23491..23791,23882..24020,24112..24381,24466..24552,
24646..24934))
/gene="F7A19.6"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39283.1"
/db_xref="GI:5080773"
/translation="MGGCVSQRKLSNKKLOOKKRRGRSGCKRSKISASHPDVPMKRA
SNASVPAVSEVYHLDPNCAAKMKRKBMNANHLLTOLWNCSQIDGNRISHBE
AWTQSFSTIDSDSGSNSSVFEEDANASAMGOVLYQEFSEYSLKIDSGNETYSNA
EVSIRKNOVDSDHHEHFKTTTCDHDHRRKSSKVYVAVSVRTSIDSGSTPSESSA
EKLLYRACAGSIORSLGEMTSGQSMSEVSFSLKRLGFEFDRKCKCAPNCSPYI
PIGVDFACPKIKINHIOHLEIPLKRPASSVOYDIPMLILVNIOLDPYPSMPCDYG
EGSLVLYFKRNNENYHKEISHFKETTKRHEDEMEKVGKPTRESTIVPFRERLKIMAG
```

```
Query Match      41.8%; Score 695.2; DB 8; Length 112126;
Best Local Similarity 70.0%; Pred. No. 1,36-166;
Matches 988; Conservative 0; Mismatches 403; Indels 21; Gaps 3;

QY 268 GATGAGCTTCGCGAGGGCTACTGCTTCGTGGTTTGATGAAAGATCTTGGCTAGTAGG 327
DB 61766 GATGAGCTTCGTAGGAGGACTTTTAAACCGCTGATTTGATGAAAGATCTTGGCTAGTAGG 61825
QY 328 TACCAATGATTCATTACCGTAAACCTTCACCTTACAAGCCATCTTCTATCTCATCTCT 387
DB 61826 TATCACTGCTTCCTTATTCGCAAGCCTTCGGCTTACAACAACCTTCGATATCATCTCT 61885
QY 388 AAGCTTAACAACTACGAAGACTTCACAAGCGATGTGGTCCGGGTACTGAATCTTACAAG 447
DB 61886 AAGCTTAACAACTATGAGATGCTCCACAAGCGTTCGCGCTCCAGCAGTGAAGTACAAG 61945
QY 448 AAGGCTTAACAACTGATCATCAAGAACATATGATG-----GTGATGGTGA 495
DB 61946 AGAGCAACAGAGAAAGCTTGGTCATGATCATGATGATGATGATGATGATGATGATGATG 62005
QY 496 TGCAAAATATGTTGTGATGATTTCTTTAGCGGCTTAGGGAACAGAAATCTTCTCTAGCC 555
DB 62006 TGCAAAATCATTTGTGTGATGATTTATGCGCTTGAAGAAATACCTTACTCTGCT 62065
QY 556 TCGGTTTCTTTCGCGGCTTTTAAGGATATGATGATGATGATGATGATGATGATGATGATG 615
DB 62066 TCTGTGTTCTCTACGCTCTCTTACAGAGAAATCATCTTGTGAGCAACAGCAAGCAT 62125
QY 616 ATGATGATCTCTTTTGGAGCCGCTTCTCGTATGTCGTGTTGATGATGATGATGATGATG 675
DB 62126 ATGATGATCTCTTCTGAGACCTTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCG 62185
QY 676 CCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 735
DB 62186 CCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 795
QY 736 GTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 795
DB 62246 TTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 62305
QY 796 CATGATATGAGATCATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 855
DB 62306 CATGATATGAGATCATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 62365
QY 856 AAATGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 915
DB 62366 AAATGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 62425
```

[illegible]

REFERENCE	JOURNAL	J Biol. Chem.	275 (20),	15082-15089 (2000)
PUBMED	20270214			
AUTHORS	Zeng, W. and Keegstra, K.			
TITLE	Direct Submission			
JOURNAL	Submitted (11-JAN-2000)			
	State University, Room 110, Plant Biology Building, East Lansing, MI 48824, USA			
FEATURES	Location/Qualifiers			
Source	1. 1698			
	/organism="Pisum sativum"			
	/db_xref="taxon:3888"			
CDS	1. 1698			
	/function="involved in cell wall xyloglucan biosynthesis"			
	/note="alpha (1,2) fucosyltransferase"			
	/codon_start=1			
	/product="xyloglucan fucosyltransferase"			
	/protein_id="AAF62896.1"			
	/db_xref="GI:7453579"			
	/translation="MNNLIKRVIAIKNPRGDDNNNKLSDLETLTDRKCTCPPLTMRV MAFVPSMFLSVPLSLSVYRDRPSPASISPTLLRNLGSLDSDPSEVLLNDKTL LGGLLADGFDEKSLSRQSAIFPKGLSIRKRYERHHQCCGYETTSYKTK TVKLGLSGQSESESVDCRYWVIFSGJGNRLTLVSAFLVALDLDRVLVDPGVMDT LFCEPFPDASFPDPDEPLNSLNNFNNQSNQCGKILKTSINSTNVSPLYDHLAA DYDHDKLFCEDEQLFQNPPLIMKTDNFISPLFSEQELNDLEPKKRVFPH FLGRLHPHTNNWGLVRYDAYLAKVADENIGIQLRFPDPPPOHVDLOVACTIG KESILPYNREONINSSSGPKSAVALITLSISQYFEKVRWMEPPPEPGEVGIQV PSHGQYQOTOKQFNNOKAAMAMYLILSLFDVLYTSNPSFCVYAGLGLKRWILYKEP NRINPNCORAMSEPEFTNAPPYIDDKAKRGDTGVALPVHRHCEMSWKLVDN"			
BASE COUNT	480 a 357 c 340 g 521 t			
ORIGIN				
Query Match	38.7%: Score 642.6; DB 8; Length 1698;			
Best Local Similarity	66.3%: Pred. No. 3.8e-153;			
Matches	924; Conservative 0; Mismatches 469; Indels 0; Gaps 0;			
267	TGATAAGCTTCTCGAGAGGACTTGTGCTTGTGGTTTGATGAAGATCTTGCCTTAGTAG	326		
297	TGATTAACCTGCTCGGTGGCTTCTCGTGATGATTTGACGAAATATCTTGTCTCAGCAG	356		
327	GTACCAATCAGTTCTTACCCGTAACCTTACACTTACCAAGCATCTTCTATCTATCTC	386		
357	GTATCAGTCAGCATTTTTCGGCAAGGACTATCAGGAAACCTTCTTCTTACCTTATTTTC	416		
387	TAAAGCTTGAAGAACTTCGAAAGCTTCAACAAGCGATGTGTCGGGTACTGATCTTACAA	446		
417	TAGACTGAGAAATATGAAAGCCCGACACAAACATGTGGACTTACTGATCATTTATRA	476		
447	GAAAGCTTAAACACACTTGATCAAGAACATATGATGTGATGTGAATGCAATATGT	506		
477	CAAAACGGTGAAGAAACGCTTCTGGTCAATATCTGATGCCGTGGATGTGAATATGT	536		
507	TGTGAGATTTCTTTTATGAGGCTTGAAGGACAGATCTTCTTCAAGCTCGGTTTTCT	566		
537	AGTATGATTTCTTATGAGGCTTGAAGAAATAGATATGACCTTATGACATTTCT	596		
567	TTACGCGCTTTTAAACGATAGAGTCTTCTGTTTACCGAGGAAAGACATGATGATCT	626		
597	TTACGCTCTCCATCATGATCGTGTCTTCTACTAGTTGACCGGAGATCGATGATGATCT	656		
627	CTTTTGACAGCGCTTTCGCGATATCTCGTGGTGTCTACTTTAGATTTTCCATGACGA	686		
657	TTTTTGTGAACCATTTTCTGACGCTTCCGTGTTTCTCCCTCTGATTTTCTCTTAAAG	716		
687	TCAGTTTATGATGAATTAATCAAGATTCATCTCGTTATGATATATATGATGAAGATCA	746		
717	TCATTTGAATATATTTCAACACGAGATTCAGTACGTCTACGGGAAATATACGTAACAAA	776		
747	GGTATGATATACTAGGGAATTTGTCTCATCTTATCTTATCTTGTTCATGATTAATGG	806		
777	ATCAATTCACAAATTCACATGTGCACTCTTTGTTTATCTTCATCTACGACCATGATTAACA	836		

	/gene="At2g15440"
	/codon_start=1
	/product="expressed protein"
	/protein_id="AAD2280.1"
	/db_xref="GI:4544369"
	/translation="MKSNTNTNLFIHHHPLTSKSPSAVAANHHFLLTFLESTFLHSFLPSSISHITSHSLSSPSLPPVLAALHYTSSSPNTPMSFLESTISNIHSHGPNCLILGLTHESLIHMSINFOGTIVDSDPYSVSKFEQSNPGVEYXEVYSTKVSOAKKLIGYKTRPECRPVQNILSDCKIINDLPFYVEIMDVLLIDGGPRGASDPGMARIPFTSAVLAKSODGFKTKRDVDLVHFHGRIKEVYSDEEFLCEENLIEVGDLGHFHVVAAAEERSGDGPCFRNSTRLSESTPAVSGDEEKVGDDE"
repeat_region	complement(13550..13679)
	/rpt_family="(GGA)n"
repeat_region	complement(14991..15034)
	/rpt_family="AT_rich"
repeat_region	complement(16086..16198)
	/rpt_family="AT_rich"
repeat_region	16533..16571
	/rpt_family="AT_rich"
repeat_region	complement(18723..18757)
	/rpt_family="AT_rich"
repeat_region	complement(18897..18937)
	/rpt_family="AT_rich"
repeat_region	18940..18991
	/rpt_family="(TAANA)n"
repeat_region	19685..19758
	/rpt_family="Rf:g[11100803]emb X93607.1 AT74A.A.thaliana
repeat_region	microsatellite repeat region (clone 74A)#Rf:ATR0046 X93607
	74A"
repeat_region	19685..19757
	/rpt_family="Rf:g[11100803]emb X93607.1 AT74A.A.thaliana
repeat_region	microsatellite repeat region (clone 74A)#Rf:ATR0046 X93607
	74A"
repeat_region	20089..20135
	/rpt_family="AT_rich"
repeat_region	complement(20944..20991)
	/rpt_family="AT_rich"
repeat_region	complement(21056..21089)
	/rpt_family="AT_rich"
repeat_region	21180..21230
	/rpt_family="AT_rich"
repeat_region	complement(21284..21309)
	/rpt_family="AT_rich"
repeat_region	complement(21445..21487)
Query Match	36.5% Score 607.4; DB 8; Length 89301;
Best Local Similarity	65.0%; Pred. No. 5.4e-144;
Matches 969; Conservative 0; Mismatches 501; Indels 21; Gaps 4;	
OY 175 TCAATGATCTTTTCACCAACCACCATCGATTCCAATCGATTATGGTTGCAGACT 234	
Db 47372 TCACGTTTGTGCATTCATTGTAGCTAAGAATTCCTTTTGACCTTTTTTTGT 47431	
OY 235 AGAGTTCTGACGCCGAGTTTCCCAAATTCGTATAAGCTTCGAGGCGCTACTTCT 294	
Db 47432 ATATGTTACAGCAGATCCGAACAACCGGGCGGATACATAATAGAGGGCTTTAAC 47491	
OY 295 TCTGTTTGTGAAGAATCTTTCCTTAGTAGTAC--AATCAGTTCAATTACGTAA 351	
Db 47492 GCAGATTTTCGATGAAGTTCTGCTGTGAGTATCATTAACCTTCTTATCGCAAG 47551	
OY 352 CATTACCTTCAGGCCATCTTATATCCATCTCTAGCTTAGAAGTAACCTAGAAAAGCTT 411	
Db 47552 CCTTCACCATCAACCCCTCGAATATCTTGTCTGGAAGCTTAGAGCTATAGATGCTT 47611	
OY 412 CACAAGCATGTGGCCGGTACTGAATCTTACAGAAGAAGCTCTAAACAACCTGATCAA 471	
Db 47612 CACAAGCTTGGGCTCACGGGCAAAAGCTTACAGAAGAACAACAAGCATCTTAGTAT 47671	
OY 472 GAACATATTGATGCT-----GATGGTGATGCCAATATATGTTGTGTGATTTCT 519	
Db 47672 GATGAGATTATTAATGCAAGCAAAATCATGATGTGTAATCCGATACGTTGTGGCTGCT 47731	

```
QY 520 TTACCGGCTTAGGAACAGATCTTTCTAGCCCTGGTTTTCTTACGGCCTTTA 579
Db 47732 GATTACGGGGTTGGAAACCGCATCTACTCTTCTCTTCTTCTTCTCTT 11
QY 580 ACGATAGAGTCTGCTTGTGGACCGGGAAGAAAGATGATCTCTTTGGCAGCG 639
Db 47792 ACTGATGATCATCTTCTTGTGGACACCGCAGAGATATGGTATCTCTTATGCGAGCA 47991
QY 640 TTTCTGGTATGCTGCTGCTACCTTTGATTTCCCTATGCTGATCATGTTGATGA 47851
Db 47852 TTTCCAGTACTCATGCTGCTCTCTCTGACTTTCCATGTAATATCTGATGA 47911
QY 700 TTAATCAAGATCATCTGCTGTATGATATATGTAAGATCAGTGTATGATCT 759
Db 47912 TACCACAGGATGATCTGCTGTTTACGGAACATGTTGAAATCATTCCTCAACTCG 47971
QY 760 GAGGGAATCTTCTCATCTTATCTTATCTTCTGTTCAATGATATGAGATCATGATAG 819
Db 47972 ACTTCATTCCTCCGACATCTATATATGCTAACCTTCATGATCAAGGATAGTATAG 48031
QY 820 ATGTTCTCTGAGAGAGACCAACATTCATCGGAAAGTCCCTGGTATGTTTAA 879
Db 48032 ATGTTCTCTGCAAAAGATCAAGATTTGATGACAAAGTCCCTGGTATATTTAGA 48091
QY 880 ACAGACATTTACTTGTTCATCTCTGCTGTTAATACCGGCTTGCATGATGAACTAAC 939
Db 48092 GCCATGTTTACTTGTTCATCTTATGTTTATCCAACTTCCAAACCGGATGATAG 48151
QY 940 AAGCATTTCCCAAGAAAGCACTGTCTTTCATCTAGTAGTATGCTTTTACCCCA 999
Db 48152 AAGCATTTCCCAAGAAAGCACTGTCTTTCATCTAGTAGTATGCTTTTACCCCA 999
QY 1000 ACTAACAGATATGGGGCTTATGATATGATATGATATGATATGATATGATATG 1059
Db 48212 AAAATCAAGTTTGGGATATGCTTACTATGATATGATATGATATGATATGATATG 48271
QY 1060 GAGAAGATTTGGGATCAATGATATGATATGATATGATATGATATGATATGATATG 1119
Db 48272 GAGAAGATTTGGGATCAATGATATGATATGATATGATATGATATGATATGATATG 48331
QY 1120 ATGATGATATTTGATATGATATGATATGATATGATATGATATGATATGATATG 1179
Db 48332 ATGATGATATTTGATATGATATGATATGATATGATATGATATGATATGATATG 48391
QY 1180 GTGAGAGATCTGCCATGT--TAAATCCCAACAAAGCGTGTCTGATATGATATG 1236
Db 48392 GATGAGATCTGCCATGT--TAAATCCCAACAAAGCGTGTCTGATATGATATGATATG 1296
QY 1237 TTGAACGCGGCTTATGATATGATATGATATGATATGATATGATATGATATGATATG 1296
Db 48452 TTGATCTGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 1296
QY 1297 GGAGAAATCATCTGCTTATGATATGATATGATATGATATGATATGATATGATATGATATG 1356
Db 48512 GGAGAAATCATCTGCTTATGATATGATATGATATGATATGATATGATATGATATGATATG 1356
QY 1357 ATGCAATATGCAAAAGCTTGTGCAAAATGATATGATATGATATGATATGATATGATATGATATG 1416
Db 48572 GTTCAGAGACCAAAAGCGTGTGATATGATATGATATGATATGATATGATATGATATGATATG 1463
QY 1417 ACAAGTCTGCTTATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 1476
Db 48632 GCAAGCTCAAGTCAAACTTGTGATATGATATGATATGATATGATATGATATGATATGATATG 14863
QY 1477 ATACTTATGACCGGAAACCGTATGATATGATATGATATGATATGATATGATATGATATGATATG 1536
Db 48692 TTACTTATGACCGGAAACCGTATGATATGATATGATATGATATGATATGATATGATATGATATG 1536
QY 1537 ATGAGCCTTGTTCATCTGCTTCAATTCATGATATGATATGATATGATATGATATGATATGATATG 1593
Db 48752 ATGAGCCTTGTTCATCTGCTTCAATTCATGATATGATATGATATGATATGATATGATATGATATG 1593
QY 1594 GACACGGGAACACTAGTCTCTCATGATGATATGATATGATATGATATGATATGATATGATATG 1644
```

```
Db 48812 GAATCGGGGAAGTGTCTTGTGATATGATATGATATGATATGATATGATATG 48862
RESULT 7
AX412254
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT 433 a 307 c 319 g 453 t
ORIGIN
Query Match
Best Local Similarity 36.5%; Score 606; DB 6; Length 1512;
Matches 967; Conservative 0; Mismatches 500; Indels 21; Gaps 4;
QY 178 ATGATCTTTCACCAACCCATGATCAATGATATGATATGATATGATATGATATG 237
Db 1 ATGATCTTTCATCTGATATGATATGATATGATATGATATGATATGATATGATATG 60
QY 238 GTTTCACCGCGGAGTTTCCCAATCTGATATGATATGATATGATATGATATGATATG 297
Db 61 TGTTCACCAATCCCAACACCGGCGGATGATATGATATGATATGATATGATATGATATG 120
QY 298 GTTTCATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 354
Db 121 GATTCATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 180
QY 355 TCACCTTCAAGCATCTTCTTATGATATGATATGATATGATATGATATGATATGATATG 414
Db 181 TCACCTTCAAGCATCTTCTTATGATATGATATGATATGATATGATATGATATGATATG 414
QY 415 AAGCATGCTGCGGATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 474
Db 241 AAGCATGCTGCGGATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 474
QY 475 CATATGATGAT-----GATGATATGATATGATATGATATGATATGATATGATATGATATG 522
Db 301 GATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 522
QY 523 ACGGCTTATGGAACAGGATCTTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAT 360
Db 361 TACGGGCTTGAACACGATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAT 420
QY 583 GATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAT 642
Db 421 GATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGATATG 480
QY 643 CTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAT 702
Db 481 CCAAGTCTTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
QY 703 AATCAAGATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 762
```

Dh 541 CACAAGGATACCTCTGTTGTACGGAACAATGTTGAAATCATTCATCACTGACT 600
Qy 763 GGAACCTTGTCAATCTTATCTTCATCTGTTCATGATATGAGATCATGATTAAGATG 822
Dh 601 TCATTCGCCGCAACATCTATATGATCAATCCTTCATGATCAAGGATATGATTAAGATG 660
Qy 823 TTCTTCTGTGAGAGACCAACATTCATGCGGAAAGTCCCTGGTTGATGTTGTTAAACA 882
Dh 661 TTCTTCTGCCAAGAGATCAAAAGTTGATGACAAAGTCCCTGGTTGATATTTAGAGCC 720
Qy 883 GACAAATCTGTTGCTTCCATCTGTGTTATACCGGTTTCGATGATGAAGTAACAG 942
Dh 721 AATGTTTACTTGTTCATCTTATGTTTAACTCAACTTCCAAACCGAATCAACGAG 780
Qy 943 CTATTCACCAAGAAAGCACTGTCTTTCATCATCTAGTAGTATCTTTTCACCAACT 1002
Dh 781 CTGTTCCGCAAGAAAGAACCGTGTTTCACCACTGGGTCGATCTTTTCACCTTAA 840
Qy 1003 AACCAAGTATGGGGCTTAGTCACTAGATACGAAAGCTTACTTATCCATGCGGATGAG 1062
Dh 841 AATCAAGTTGGGATATGCTCACTAGTACTACATGATCACTTATCCAAAGCAATGAG 900
Qy 1063 AAGATGGGATTCAGTAAAGATTTTCATGAAGACCGGGTCCATTCAGCATGTATG 1122
Dh 901 AGACTCGGATTCAAATTCGGGGTTTTTCGCGATCAAGGTGATACCAACACGCTCATG 960
Qy 1123 GATCAGATTTCTATCTTACTCAAAAAGAAACTTCACTACGATGAGACACTAGT 1182
Dh 961 GACCAAGTCATATCTCTGCACACAAAGAAACTTTGCTGATTTGGCTTACACAGAA 1020
Qy 1183 GAGAGATCTGCCCATGT--TAATACCCCAACAAACGCGGCTTGTCCATCTTTG 1239
Dh 1021 GAATCAAAAGTCAACATATCAACATCTCCAAAGTAAGAGCTTGTCTGTACCTTTTG 1080
Qy 1240 AACGGGGTTACGGCGAAGAACTTAAAGATATGTTGGGATATCCGACATCACTG 1299
Dh 1081 TCTCTGATCTCTAAGAACTAGAGAACTGTTTTCAGACGAGCAAAATATGACGGGA 1140
Qy 1300 GAATCATACGGGTTCATCAGCCGACGCAAGAGGTATTCAGACAGCGAAAGAAAGATG 1359
Dh 1141 GAGATATTAAGATATATTCAGCAAGTGAAGAAAGTATCAACAAACAGACAAAGAT 1200
Qy 1360 CATATGCAAAAGCTCTTGGCGAAATGATCTTTTGAAGTTGACAGATATCTTGACA 1419
Dh 1201 CACGACCAAAAGGCGCTTGCAGATGTATCTTTGAGCTTAACCGATTAACATTTGTCCA 1260
Qy 1420 AGTGTGTCTACATTTGATATGATGATGATCAAGGTCTTGAAGGTTTAAAGCTTGAT 1479
Dh 1261 AGCTCAAGGTCAACATTTGATATGATGATGATGATGATGATGATGATGATGAT 1320
Qy 1480 CTCTATAGACCCGAAACCGTACACATCTCCGATCCTGCTGCTGCGGCTATGTCATG 1539
Dh 1321 CTATATCTGCAAAATGATATACAAAGCTCTGATCGCGCTGTGTCATCCAGCTGATG 1380
Qy 1540 GAGCCTTGTTCACCTGCTCCATCTATGATATGATTAAGCGAA--AACGGTATTTGAC 1596
Dh 1381 GAGCCTTGTTCCTTACTCTCCGACATGATGATGATGATGATGATGATGATGATG 1440
Qy 1597 ACGGAAACATAGTTCTTCATATGATGATGATGATGATGATGATGATGATGATG 1644
Dh 1441 TCGGGGAAAGTTGCTTTGTTAGGATGATGATGATGATGATGATGATGATGATG 1488

RESULT 8
AX412452 1512 bp DNA linear PAT 14-JUN-2002
LOCUS AX412452
DEFINITION Sequence 216 from Patent WO0222675.
ACCESSION AX412452
VERSION AX412452.1 GI:21444910
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
AUTHORS Glazebrook, J., Wang, X., Dangl, J. L., Eulgem, T. and Zhu, T.
TITLE Plant genes, the expression of which are altered by pathogen infection
JOURNAL Patent: WO 0222675-A 216 21-MAR-2002;
SYNOPSIS Syngenta Participations AG (CH); UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL (US); Glazebrook, Jan (US); Wang, Xun (US); Dangl, Jeffrey L. (US); Eulgem, Thomas (US)
FEATURES
source Location/Qualifiers
1.1512
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
BASE COUNT 433 a 307 c 319 g 453 t
ORIGIN
Query Match 36.5%; Score 606; DB 6; Length 1512;
Best Local Similarity 65.0%; Pred. No. 8.5e-144;
Matches 967; Conservative 0; Mismatches 500; Indels 21; Gaps 4;
Qy 178 ATGATCTTTCACCAACACCATCTGATTCATATCGGATTTATGGTTTCGCCGACCTAGA 237
Dh 1 AATGTTTGTCAATTCATTTGATTCACAAAGATTCCTTTGACCTTTTGTATATA 60
Qy 238 GTTTCGACGCGGAGTTTTCACAAATTCGTATAGCTTCTCGAGGCTACTTCTCT 297
Dh 61 TGTTCAGACGAATCCGAAACACGGGGGGGATAGACTATATAGAGGCTTTTAAACCGCA 120
Qy 298 GGTTTTGAATGAATTCCTGCTTATAGTATGAC--AATCAGTTTCAATACCGTAAACCT 354
Dh 121 GATTTTCGATGAAGGTTCTGCTTGTAGTATGATCAATAAATCTTCTTGTATGCAAGCT 180
Qy 355 TCACCTTCAACCCATCTTCTATCTCATCTTATAGCTTATAGAACTAGCAAAAGCTTAC 414
Dh 181 TCACCATCAACGCGTGAATATCTTGTCTGAAGCTTAGAAGCTATGAGATGCTTAC 240
Qy 415 AAGCATGTGTCGCGGACTGATCTTACAGAAAGCTTAAACAACTTATGATCAAGA 474
Dh 241 AAAGCTTGGGTCGAGGCAAAAGCTTACAGAAAGCAACAAAGCTTATGTCATGAT 300
Qy 475 CATATTTAGTGT-----GATGATGATCAATATGTTGTGTGATTTCTTTT 522
Dh 301 GAGAAATTTAATGCAAGCAAAATCAGATGATGATGATGATGATGATGATGATGAT 360
Qy 523 AGCGGCTTAGGCAACGACTACTACTCTGCTGCTGCTTCTTACGCGCTTTAACG 582
Dh 361 TAGGGCTTGAACCGACTACTACTCTGCTGCTGCTTCTTACGCTTCTTGACT 420
Qy 583 GATAGAGTCTGCTTGTGACCGAGGAAAGATGATGATCTTTTTCGAGCCGTTT 642
Dh 421 GATATATATATCTTGTGACAAACCGCAAGATATTTGGATCTTATGACGACCATTT 480
Qy 643 CTCGGTATGTCGTTGCTGCTTATGATTTCCATGACATGATCAGTTTATGATGATTA 702
Dh 481 CCAGGTACTTCAATGATGCTTCTTCTGCTTCCATGATGATGATGATGATGATGATG 540
Qy 703 AATCAAGATCATCTCTGTTGATGATATATGAGAGATCAGTATGATGATGATGATG 762
Dh 541 CACAAGGATCTCTGTTTACGGAACAATGTTGAAACAACTTCAATCAACCTGACT 600
Qy 763 GGAACCTTGTCAATCTTATCTTATCTTCTGATGATGATGATGATGATGATGATGAT 822
Dh 601 TCATTCGCCGCAACATCTATATGATCAATCCTTCATGATCAAGGATATGATTAAGATG 660
Qy 823 TTCTTCTGTGAGAGACCAACATTCATCGGAAAGTCCCTGGTTGATGTTGTTAAACA 882
Dh 661 TTCTTCTGCCAAGAGATCAAAAGTTGATGACAAAGTCCCTGGTTGATATTTAGAGCC 720
Qy 883 GACAAATCTTGTTCATCTGTGTTATACCGGTTTTCGATGATGAAGTAACAG 942
Dh 721 AATGTTTACTTGTTCATCTTATGTTTAACTCAACTTCCAAACCGAATCAACGAG 780

FEATURES	source
LOCUS	AX412825
DEFINITION	Sequence 589 from Patent WO0222675.
ACCESSION	AX412825
VERSION	AX412825.1
KEYWORDS	GI:21445283
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS	Glazebrook, J., Wang, X., Dangl, J. L., Eulgem, T. and Zhu, T.
TITLE	Plant genes, the expression of which are altered by pathogen infection
JOURNAL	Patent: WO 0222675-A 589 21-MAR-2002; SynGenea Participations AG (CH) ; UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL (US) ; Glazebrook, Jan (US) ; Wang, Xun (US) ; Dangl, Jeffrey L. (US) ; Eulgem, Thomas (US)
FEATURES	Location/Qualifiers
source	1..1512
	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"

BASE COUNT	433 a	307 c	319 g	453 t	ORIGIN
Query Match	36.5%;	Score 606;	DB 6;	Length 1512;	
Best Local Similarity	65.0%;	Pred. No. 8.5e-144;			
Matches 967;	Conservative	0;	Mismatches 500;	Indels 21;	Gaps 4;
OY 178	ATGATCTTTCACCAACACCCACTGCTGATTCAAATCGATTTATGGCTTTGCCCAAGCTAGA	237			
Db 1	ATGTTTGTGATTCATGTTAGCTCAAGAAATTCCTTTTACCTTTTATTTTATATA	60			
OY 238	GTTCGTGACCGCGGAGCTTTTCCCAATTCGTATTAAGCTCTCGAGAGGCTACTTCTCT	297			
Db 61	TGTTACAGCAAGAACCAACACCGGGCGGGATAGACTATATAGAGAGGCTTTTAACCGCA	120			
OY 298	GGTTTGTGATGAAGATTCCTTGCTTAGTAGTACC--AATCAGTTCAATTACCGTAACCT	354			
Db 121	GATTTGATGAAGGTTCTTGCTGATGATGATGATATCAATAAATCTTCTGATGCAAGCT	180			
OY 355	TCACCTTACAGCATCTTCTTATGCACTCTGTAAGCTTAGAAGTAAGTACCAAGAGCTGAC	414			
Db 181	TCACCTTACAGAGCGCTGTAATATCTTGTCTGGAAGCTTAGAAGCTATGAGATGCTTAC	240			
OY 415	AAGCATGTGTCGGGTACTGATTCACAAGAAAGCTCTAAACCACTGATCAAGAA	474			
Db 241	AAACGTTGGCGTCCAGGGACAAACCTTACAGGAAGCAACAAACATCTTAGTCATGAT	300			
OY 475	CATTTGATGGT-----GATGGTGAATGCCAATATGTTGTGGATTTCTTT	522			
Db 301	GAGATTTAATTCAGCAAGCAATCAATGATGATGCCGATACGTTGTGGCTGCTGAT	360			
OY 523	AGCGGCTTAGGAGACAGATACTTCTCTAGCGCTGGTTTTCTTTACGCGCTTTAACG	582			
Db 361	TACGGGCTTGAACCGCACTACTACCTCTGCTTGTGTTCTGATGCTCTGACT	420			
OY 583	GATPAGCTGCTGTTGTGACCGAGGAGAAAGCATNGATCTCTTTTGGAGCGCTTT	642			
Db 421	GATPAGATCATCTTGTGTGACCAACCGCAAGGATATGGTGTATCTTTTGGAGCGCATTT	480			
OY 643	CTGGGTATGCTGTTGCTACTACTTTTAATTTCCCTATAGCATGATCACTTGTATGATTA	702			
Db 481	CCAGTACTTCACTGTTGCTCTCTCTGATGCTTCCATTTGATGAATAATATCGATNGATAC	540			
OY 703	AATCAAGATATCTTCGTTGTTATGATATATGTTGAAGAAATCAGTGATTTGATCTGAG	762			
Db 541	CACAAGGATACTCGGTGTGACGAACAATGTTGGAATAATCAATCCATCACTGCACT	600			
OY 763	GGAACTTGTGCTACTTATCTTATCTTCACTTGTGATGTTTGGATGATTAAGATAG	822			
Db 601	TCATTCGGGCACATCATATATATGATTAACCTTATGATTAAGAGATAGATTAAGTG	660			
OY 823	TTCCTCTGAGAGGACCAAAATTCATCGGGGAAAGTCCCTGGTGTATGTTAAACA	882			
Db 661	TTCCTCTCCAAAAGATCAAAAGTTTGATTTACAAAGCCCTTGATATTTATAGACC	720			
OY 883	GACATTAATCTTGTCCATCTGTGTTGGTTATPACCGGGTTTCGATNGATGAATCAACAG	942			
Db 721	AATGTTTACTTTGTTCCATCTTTATGTTTATCCAACTTTCAAACCGAATACGAAAG	780			
OY 943	CTATTCCACAGAAAGGAGCTGTCTTTCATCACTTAGTAGGATTTCTTTTACACCCA	1002			
Db 781	CTGTTTCCGCAAGAAAGAACCGTCTTTCACCACTTGGGTGGTATCTTTTACACCTAA	840			
OY 1003	AACCAAGTATGGGCTTAGTACATAGATTAACAGACGTTTACTTATGATGCGGATGAG	1062			
Db 841	AATCAAGTTTGGGATATGCTACATAGATTAACATATATCATCTTATCCAAAGAGATGAG	900			
OY 1063	AAGATTGGATTCAGTAAGAAGTTTGTGATGAGAACCGGGTCCATTTACGATGATG	1122			
Db 901	AGACGTGGGATTCAAATTTGGGTTTTTCGGGATGCAAGGTGATCTACCAACAGCTGATG	960			
OY 1123	GATGAGATTTTACTTGTACTTCAAAAAAGAAACTTTTACCTGAGTAGACACACTAGTG	1182			

Db 961 GACCAGGTCATATCTCGACACACAAAGAGAAACTTTTGCTGATGTTGGCTACACAGAA 1020
Qy 1183 GAGAGATCTGGCAGTGT---TAATACCCCAACCAAGCCGCTGTGCATCTTTG 1239
Db 1021 GAATCAAAAGTCACATATCGAACATTTCCAAAAGCTAAAGCTGTTCTGTCTCACTTTTG 1080
Qy 1240 AACGGGGTTCAGCGGAGAACTTAAAGATGATGTTGGAATATCCGACATCACTGGA 1299
Db 1081 TCTCTGATGTAAGTAAGAACTAGAGAAATGTTTTCAGAGCCGACAAATATGACGGGA 1140
Qy 1300 GAATTCATCGGTTCATCAGCCGACCAAGAGGTATTCAGACAGCCGAAAAAAGATG 1359
Db 1141 GAGATTATTAAGATATATACCAAGCTCAAGTATCAACAAACAGACAGAGAGTT 1200
Qy 1360 CATATGCGCAAGCTCTTCCGGAATGTATCTTTGATTTGACATGAATCTTGACGA 1419
Db 1201 CACGACCAAAAGCCGCTGCTGAGATGATATCTTTGACCTTACCGATTAACATTTGTGA 1260
Qy 1420 AGTGGTGTGTACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1479
Db 1261 AGCTCAAGGTCAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
Qy 1480 CTCTATACACCGAAACCGATCAACTCCGATCTTGTGCTGGGCTATGCTGATG 1539
Db 1321 CTATATCTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
Qy 1540 GAGCTTGTTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1596
Db 1381 GAGCTTGTTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Qy 1597 ACGGACACATAGTTCCTCATGTGAGACATGTTGAGATGATGATGATGATGATGATGATG 1644
Db 1441 TCGGGGAAGTGTCTCTTTGTTGATGATGATGATGATGATGATGATGATGATGATG 1488

RESULT 10
AF417474 1512 bp mRNA linear PLN 27-DEC-2001
LOCUS Arabidopsis thaliana fucosyltransferase-like protein FUT4 (FUT4)
DEFINITION mRNA, complete cds.
ACCESSION AF417474
VERSION AF417474.1 GI:17980429
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE 1 (bases 1 to 1512)
AUTHORS Sarría, R., Wagner, T.A., O'Neill, M.A., Faik, A., Wilkerson, C.G., Keegstra, K. and Raikhel, N.V.
TITLE Characterization of a family of Arabidopsis genes related to xyloglucan fucosyltransferase.
JOURNAL Plant Physiol. 127 (4), 1595-1606 (2001)
MEDLINE 11743104
PUBMED 21608393
REFERENCE 2 (bases 1 to 1512)
AUTHORS Sarría, R., Keegstra, K. and Raikhel, N.V.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2001) PRL, Michigan State University, 122 Plant Biology Building, East Lansing, MI 48824, USA
FEATURES
SOURCE Location/Qualifiers
1..1512
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="2"
/map="At2g15390"
1..1512
/gene="FUT4"
1..1512
/gene="FUT4"
/note="AtFUT4"
/codon_start=1

/product="fucosyltransferase-like protein FUT4"
/protein_id="AAL50623.1"
/db_xref="GI:17980421"
/translation="MCHSLAQRIILLTFEYVCSDBSETPGDRILGILLTDPDFE
GSLSKRYHKEFLYKRPSPKPESELYLVKRELYEMLHKRCGCTPAVKEAVKRLSHDEN
YNASKSDGECEYVVMADYGLGNRLTLASVFLVALLDRILLVNDNRKIDGLDCEFP
PGTSMLEPLDEPLMKRYADYKRSYRCGYTLEHNSINSTSEPHLVMHNLHDSRSD
KMFECODOSLIDKVPMLIFRANYVPEPMLFNPOTELKLEPKEKTEVPHHNGRYL
FHPKNOYMDIYTKYVHDLHLSKADRLGILFVPEPDGQYVHVMDOYISCTQREKLLP
ELATOESKVINISIPKSAVLTVLSVSTPERSKLENNSEKANKTGEITIKYOPSGER
YQOTDKRVHDKALAEMLVLSLTDNISVSSSTGYVAYSLGGLKPMILVLPNNKAP
DPPCVRSTSMPCPLTPPTHECEDPDAMGTESGKVPVPRVCEIDIMGLDEDEL"
BASE COUNT 433 a 307 c 319 g 453 t
ORIGIN
Query Match 36.5%; Score 606; DB 8; Length 1512;
Best Local Similarity 65.0%; Pred. No. 8.5e-144;
Matches 967; Conservative 0; Mismatches 500; Indels 21; Gaps 4;
Qy 178 ATGATCTTTACCAACACCATCTGATTCGAATTCGATATATGATGATGATGATGATGATG 237
Db 1 ATGTTTGTCTATTCATTTGTTAGCTCAAGAATTTCTCTTTTACCTTTTGTATTA 60
Qy 238 GTTCTGACGCGCGGAGTTTCCCAATTCGTATGAGCTTCTCGAGGCTACTTCTCT 297
Db 61 TGTTCAGAGCAATCCGAACACCGGGGGGATACACTAATGAGAGGCTTTAAACCGGA 120
Qy 298 GGTTCATGAGAAATCTTGCTTGTAGTACG---AATCAGTTTCAATACCGTAACT 354
Db 121 GATTTGATGAAAGGTTTGTGCTGATGATGATGATGATGATGATGATGATGATGATG 180
Qy 355 TCACCTTCAAGCCATCTCTTATCTCATCTAGCTTAGCAATACGAAGAAGCTTCC 414
Db 181 TCACCTTCAAGCCATCTCTTATCTCATCTAGCTTAGCAATACGAAGAAGCTTCC 240
Qy 415 AAGCATGTGTCGCGGCTACTGATCTTACAAAGAGCTTAAACAACTGATCAAGA 474
Db 241 AAACCTGGGTCGAGGACCAAAAGCTTACAGAGACAGCAACAGATCTTGATGAT 300
Qy 475 CATATTGATGCT-----GATGCTGATGCAAAATATGTTGTGATGATTTCTTT 522
Db 301 GAGAAATTAATGCAAGCAAAATCAAGATGATGATGATGATGATGATGATGATGATG 360
Qy 523 AGCGCTTAGGAGACAGATATCTTCTAGCCGCTGGTTTCTTACGCGCTTTAAG 582
Db 361 TAGGGCTTGAACCGACTACTACTCTGCTTGTGTTCTCTACGCTCTTTGACT 420
Qy 583 GATAGAGCTTCTGTTGACCGAGGAAAGACATGATGATCTTTGGAGCGCTTT 642
Db 421 GATTAATCATCTTGTGACACGCAAGATGATGATGATGATGATGATGATGATGATGATG 480
Qy 643 CTCGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 702
Db 481 CCAGCTACTTCAVGGTGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Qy 703 AATCAAGATATATCTGTTATGATATATATGATGATGATGATGATGATGATGATGATG 762
Db 541 CACAAGGATACTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Qy 763 GGAATCTTGTCACTTATCTTATCTTGTGATGATGATGATGATGATGATGATGATGATG 822
Db 601 TCATTCGCGGCACATCTATATGATGATGATGATGATGATGATGATGATGATGATGATG 660
Qy 823 TTTCTGTGGAAGACCAACATTCATGCGGAAGTCCCTGCTGATGATGATGATGATGATG 882
Db 661 TTTCTGTGGAAGACCAACATTCATGCGGAAGTCCCTGCTGATGATGATGATGATGATG 720
Qy 883 GACATTAATCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 942
Db 721 AATGTTTACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
Qy 943 CTATTCACCAAGAGAGCTGCTTTCATCACTAGGATGATGATGATGATGATGATGATGATG 1002

Db 781 CTGTTCCGCAAGAAACCGGTTCACCACTTGGGTGCGTATCTTTTCACCCCTAAA 840
 QY 1003 AACCAAGATATGGGGCTTAGTCACTAGATTAACGAGAGCTTATGCGATGGGATGAG 1062
 Db 841 AATCAAGTTTGGATATGCTTAACTAGTACCTATGCTTATATCAAAAGCAGATGAG 900
 QY 1063 AAGATTGGGATTCAGATAGAGTTTTCGANGAGACCCGGGTCCATTTTCAGCATGTGATG 1122
 Db 901 AGACTCGGATTCAAATTCGGGTTTTTCGGCATGAGGTGATGATCTCAACACGCTCATG 960
 QY 1123 GATGAGATTCATCTGTACTCAAAAAGAAACTTCTACTGATGATGACACACTGATG 1182
 Db 961 GACCAAGCTATATCTTCACACAAAGAGAAACTTTTGGCTGAATTCGCTACACAGAA 1020
 QY 1183 GAGAGATCTCGCCATGT---TAATACCCCAACACAAACCGGCTTGTGCATCTTTG 1239
 Db 1021 GAATCAAAAGTCAACATATGGAACATTCAAAAGTAAAGCTGTCTTGTGACATCTTTG 1080
 QY 1240 AAGCGGGTTACGGGGGAACTTAAGAGTATGTTGGGAATATCCGACATCACTGA 1299
 Db 1081 TCTCTAGTACTCTAAGAACTAGAGAACATGTTTTCAGAGCGAGCAAAATATGACGGGA 1140
 QY 1300 GAATCATCGGTGTTCATTCACGCGCAAGAGGTTATGACGAGCCGCAAAAAGATG 1359
 Db 1141 GAGATTTTTAAAGTATATCAAGCAAGTGGAGAAAGTATCAACAAACAGCAAGAGGTT 1200
 QY 1360 CATATGCGAAAGCTCTTGGGAAATGTATCTTTTGATTTGACAGATATCTTGTGACA 1419
 Db 1201 CACGACCAAAAGCGCTTGTGATGATGTATCTTTTGAGCTTAACGATATCATGTTGCA 1260
 QY 1420 AGGCTTGGTCTACATTTGGATATGATGCTCAAGGCTTTGGAGCTTTAAAGCTTGATA 1479
 Db 1261 AGCTCAAGGCAACATTTGGATATGTTGCTTATAGCTTTGAGAGATTAAAGCATGTTA 1320
 QY 1480 CTCTATAGACCCGAAACCTTACACTCCGATCTTCGGTGGGGGTATGCTGATG 1539
 Db 1321 CTTTATCTGCAATGATATCAAGAGCTCCGATCCGCGTGTGATGATCCAGCTGATG 1380
 QY 1540 GAGCTTGTTCACATGCGCTCCATCTTATGATTTGTAAGCGAA---AAGGGTATTTGAC 1596
 Db 1381 GAGCTTGTTCCTTACTCTCCGACTCATGATGTGAACCTGACCATGAGGGAATGAA 1440
 QY 1597 ACGGGAACCTAGTCTCATGTGACATTTGAGAGATATCACTG 1644
 Db 1441 TCGGGGAAGTGTCTTTTGTAGTATGTGAGGATATATG 1488
 RESULT 11
 AF417475 1602 bp mRNA linear PLN 27-DEC-2001
 LOCUS Arabidopsis thaliana fucosyltransferase-1-like protein FUT5 (FUT5)
 DEFINITION
 AF417475
 mRNA, complete cds.
 AF417475
 ACCESSION
 AF417475.1 GI:17980422
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1602)
 Sarrila,R., Wagner,T.A., O'Neill,M.A., Paik,A., Wilkerson,C.G.,
 Keegstra,K. and Raikhel,N.V.
 Characterization of a family of Arabidopsis genes related to
 xyloglucan fucosyltransferase1
 Plant Physiol. 127 (4), 1595-1606 (2001)
 JOURNAL
 MEDLINE
 PUBMED
 11743104
 2 (bases 1 to 1602)
 Sarrila,R., Keegstra,K. and Raikhel,N.V.
 Direct Submission
 Submitted (07-SEP-2001) PRL, Michigan State University, 122 Plant
 Biology Building, East Lansing, MI 48824, USA

FEATURES
 source location/Qualifiers
 1..1602
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="AL2915370"
 1..1602
 /gene="FUT5"
 55..1602
 /gene="FUT5"
 /note="AtFUT5"
 /codon_start=1
 /product="fucosyltransferase-1-like protein FUT5"
 /protein_id="AA150624.1"
 /db_xref="GI:17980423"
 /translation="MKVLIAVFSGLFILTSSNNFNKLLIDATKVIDKTEKPVK
 LIGLITLDPDEGSCLSRYHKYFLRKPSPKPSSEYLVSKRSEMLKRCGPTEY
 KEAIEKLSRDSSENGECRIIVWAGGNGNLTLLASVFLALFLERILINDNRD
 VSDLCFEFPQTSMLPLDPLMLNVTYAMCGNKRKPRPCGMSKSHINSTIPHLIY
 MHNLDSDSKPLFCOKDOSLIDKRVMLVQANVTVPSLMEPFTQTELVKLEPK
 ETVHHRLRIYFHPNEDWMDVDTYHAHLSKADERLGIQIRVEKPDGRFKHIDV
 ISCTQREKLLEFATPEESKYNIKTRKLSVLAISYPERGNTLTMPSKRSSTGE
 IVEVYQPSGERVQDTDKSHDOKALAEMLILSDINIVTSARSIFGVYSLSGLKPM
 LLYQPTNFTTVPNPPCVRSKSMPECLTLPSSHCEADMGTSNGRIILFVRCEDLIYG
 LKIVDF"
 BASE COUNT 481 a 331 c 321 g 469 t
 ORIGIN
 Query Match 36.3%; Score 604; DB 8; Length 1602;
 Best Local Similarity 66.6%; Pred. No. 2.8e-143;
 Matches 929; Conservative 0; Mismatches 450; Indels 15; Gaps 4;
 QY 268 GATGAGCTTTCGGAGGGCTACTCTCTGCTTTGATGAAGATTCCTGCTTAGTAG 327
 Db 181 GATTAAGCTTAAGAGGGCTTTTAACTCGGATTTGATGAAGGTTCTGCTTAGTAG 240
 QY 328 TACC---AATGAGTTCATTACCGTAAACCTTCACCTTACAGCCATCTCTATCTCATC 384
 Db 241 TATCAATAATATTTCTTGTACCGGACGATCCCGTACAAAGCTTCTGATATCTAGTC 300
 QY 385 TCTAAGCTTAGAAGCTACGAAGAAGCTTCACAAGGATGCTCGGGTACTGAATCTTAC 444
 Db 301 TCTAAGCTTCAGAGCTATGAGATGCTTCACAAGCTTGCTGCTCCAGATACGAATATTAC 360
 QY 445 AAGAAGCTTAAACAACTGATCAGAAACAT---ATTGATGCTGATGGTGAATGCAAA 501
 Db 361 AAGAAGCAATAGAGAACTTAGTCTGATGATGATGCAAGCGCAATCAATGATGAATGCA 420
 QY 502 TATGTTGTGATTTCTTTTACCGGCTTAGGAGAACAGATACTTCTCTAGAGCTCGGTT 561
 Db 421 TACATTTGATGGTGGGACGTTTACGGGCTTGGAACAGATTACTTCTTCTGCTT 480
 QY 562 TTTCCTTACGGGCTTTTACGAGATAGAGTCTTCTTGTGACGAGGGAAGACATGAT 621
 Db 481 TTCCCTAGCGCTCTTGTGACGAGAGATCATCTTGTTCACAAACCCAGAGATGTTAGT 540
 QY 622 GATCTCTTTGAGCGCGTTCCTCGTATGCGTGGTGTGACTCTTAGATTTCCCTGAT 681
 Db 541 GATCTCTTATCGAGCATTTCCAGGATCTTCAATGATGCTTCCGCTTGACTTCCATG 600
 QY 682 ACTGATCATGTTGAT-----GGATTAATCAAGATCATCTGCTGTGTATGATATATG 735
 Db 601 CTGATTAATTAATCTTATGCTTGGGCTTACAAATAGGAATATCTCTGTTTACGAGACAT 660
 QY 736 GTGAAGATCAGGATGATGATGATGAGGAACTTTGCTATCTTATCTTATCTTGT 795
 Db 661 TCTGAAGAAATTCACCACTGCACTTCAATCCCGCGCATCTATACATGATTAACCTT 720
 QY 796 CATGATATGAGATCATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 855
 Db 721 CATGATTCAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 QY 856 AAAGTCCCTTGGTGTGATTTAAACAGACATTAATCTTGTTCATCTCTGTTGTTAATA 915

||||| 781 AAAGTCCCATGTTGATGTTCAAGCCAAATGTTACTTTGTCATCGTATGTTAAAT 840
OY 916 CCGGGTTCATGATGATGACAAACAAACAACTATCTCCACAGAAAGGACTGCTTTCATCAC 975
Db 841 CCAACTTCCAAACCGAAGCTAGTTAACTGTTCCCGAGAAAGAAACCGTTTTCACAC 900
OY 976 TTAGTGTAGTATCTTTTTCACCCCACTAACCAATGATGGGGCTAGTCACTAGATCTAC 1035
Db 901 TTGGCTCGGATATCTTTTTCACCCCACTAACCAATGATGGGGCTAGTCACTAGATCTAC 960
OY 1036 GAAGCTTACTTATTCGATGCGGATGAGAGATGGAATTCAGTAAAGATTTCGATGAA 1095
Db 961 CAGCTCATTTGTCGAAAGCGGAGAGAGACTCGGATTCAAATAGAGGTTTTCGGCAA 1020
OY 1096 GACCCGGGTCATTTTCAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 1155
Db 1021 CTGATGAGGCTTTCAAAACATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
OY 1156 CTTCCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1212
Db 1081 CTGTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
OY 1213 CACAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1272
Db 1141 CTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
OY 1273 TATTGGGAATATCCGATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1332
Db 1201 TTTTCAAGGAGCCAAAGTCAAGGAGAAATGTTAGATTTTCAACCAAGTGAGAGAG 1260
OY 1333 GGTATCAGCAGACCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1392
Db 1261 AGAGTTACGACCAACAGCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGAT 1320
OY 1393 TTGATTTGACAGATTAATCTTGTGACAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1452
Db 1321 TTGAGCTTAAACGATTAATCTTGTGACAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
OY 1453 GGTCTGAGGTTTAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1512
Db 1381 AGCTTTGAGAGATTAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
OY 1513 CTTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1572
Db 1441 CCGCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
OY 1573 TGTAAAGCGAAAGCGGTATTTGACACGCGGAACACTAGTTCCTCATGTGAGACATTTGAG 1632
Db 1501 TGTGAAGCTGACTGCGGAACTACTCGGGAAGATCTCTCTTTTGTAGGCAATTTGAG 1560
OY 1633 GATTCACGTCGGG 1646
Db 1561 GATCTTATATATG 1574

RESULT 12
AP004817/c 160096 bp DNA linear HTG 07-MAR-2002
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) chromosome 2 clone P0539D10,
*** SEQUENCING IN PROGRESS ***, in ordered pieces.
AP004817
VERSION
AP004817.1 GI:19223844
ACCESSION
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:P0539D10.
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.
REFERENCE
1
AUTHORS
Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, PAC

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
OY 268
Db 99646
OY 328
Db 99586
OY 388
Db 99526
OY 448
Db 99466
OY 508
Db 99406
OY 568
Db 99346
OY 628
Db 99286
OY 688
Db 99229
OY 748
Db 99169
OY 808
Db 99115
clone:P0539D10
Published Only in Database (2002)
2 (bases 1 to 160096)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (06-MAR-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rnp.dna.affrc.go.jp/,
tel:81-298-38-7441, fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 160096
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="P0539D10"
40077 a 39091 c 39153 g 41624 t 151 others
33.4%: Score 554.6; DB 2: Length 160096;
64.0%: Pred. No. 1e-130;
890: Conservative 0; Mismatches 489; Indels 12; Gaps 3;
GATAAGCTTCTCGAGGCTACTGCTTCTGTTGATGAGAAATCTTGCCTTAGAG 327
GACCAACTACTGCGGTGTTGTTTAACTACTGATGATGATGATGATGATGATGATGATGATGATGAT 99587
TACCAATCAGTTCATTCACCAACCTTACCTTACCAAGCCATCTTCTATCCTATCT 387
TATCAATCTGATATACCGTAAGCATCATCCATTTCCATCTGCAATCTCTGAG 99527
AAGCTTGAACACTGAAAGCTTACAGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 447
AGACTAAGGGAACATGAGGCACTTCAAAAGAGTGTGCTCACAATGATGATGATGATGATGATGATGATGAT 99526
AAAGCTTAAACACTTGTATCAAGACATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 507
AAAGCTATTGAGCAGCTAAAGGCTGACAGGCGCTAAGGTTATGAGATGATGATGATGATGATGATGATGATGAT 99466
GTGTCGATTTCTTTTACGCGCTTAGGGAACGATATCTTCTAGCCTCGGTTTCTT 567
GTATGGGTAGCTTACAGTGTCTGCGGAACGATCTTGAACAATGCGCTTCAGCATTTCTCG 99406
TACGCGCTTTTAAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 568
TATGCTATCCTCACTAGAAAGATTTTACTTTCACACAGTGTAAAGCAGCTGCAATCTT 99346
TTTTGCGAGCGGTTTCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 628
TTTGTGAACCATTTCCGAAACATCATGTTTACTTACCGCCAAATATCCCTAT--TAAAG 99286
CAGTTTGAATTAATCAAGATCATCTGCTGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 688
CAGTTCAAGAAATTCAGTAAGGTTCTCCTGAGATTATGGAACATGCTGTAACAAACAAA 99229
GTGAT 748
GCTATTCCTTAAACCTGCTT-----TCTTATCTTCAATATGCTCATGATTAACAGT 99169
GATCATGATTAAGATGTTCTTGTGAAGAGACCAACATTCATGCGGAATCCCTTGG 808
GACTATGATTAAGTATTTTCTGTGAAGATTAATCAGCAATATCTTGAATATCCCATG 99115

QY 868 TTGATTTGTAACAGACATTACTTTGTCATCTCTGGTAAATACGGGTTTCGAT 927
 Db 99055 CTGATCTTAATAATTCAGATATTACTTTGACATCACTTTTCCATTCACATATCA 98996
 QY 928 GATGAACATACAGATTTTCCACAGAAAGGACTGTTTTCATCACTTAGTAGAT 987
 Db 98995 GAGGAACTCACAGGCTTTTCCCTCAGAGAGATTCGTTTCCATCACTTGGAGCTTAC 98936
 QY 988 CTTTTCACCCCACTAACAGATGATGGGCTTAGTCACAGTACTACCACTTACTTA 1047
 Db 98935 TTGTTTCAACCTCAGCAATGTTTGTGGGCAAGTCACAGATTAATGATTTCTTCA 98876
 QY 1048 TCGCATCGGAGAGAGATTTGGATTCAGTAAGATTTTGTGATGAAGCCGGGTCCA 1107
 Db 98875 GCCAGAGCTGATGAAGAGGTGGGTATCCAGATCAGAGTCTTGATCCGAAACAGACCA 98816
 QY 1108 TTTCAGCATGTATGATCAGATTTTCATCTTGTACTCAAAAAGAGAACTTCTACCTGAA 1167
 Db 98815 TTTCAGCAGCTCTTGATCAGGCTTCTTGATCAGCCTAAAGAAACCTATTTGCACGA 98756
 QY 1168 GTAGACACATGATGGAGATCTGCCATGTATATACCCCAACAGAACCCCTGCTT 1227
 Db 98755 ATCAACAGTAA---CAGCCATTTCTTCACAGAGCAATTGATTAATAATCTGTCTTA 98699
 QY 1228 GTACATCTTTGTAACGGGGTTACGCGAGAACTTAAGAGATGATTTGGGATATCCG 1287
 Db 98698 ATCACTCTCTGATTCAGATGATGATGAGAAATCAGAGATGATCTGGGAACATCA 98639
 QY 1288 ACATCACTGAGAAATATATGCGTGTATCATCAGCCGAGCCAGAAAGATTAACAGACAC 1347
 Db 98638 ACCAGCAAGGATGATATATAGCTTTCACACAGCCAGCATGAGAAACATCAAGAACTCA 98579
 QY 1348 GAAAAAAGATGCATATATGCAAGCTCTGTGGGAAATGTATCTTTAGTTTGACAGAT 1407
 Db 98578 GACAAAGAAATGCAATATGAAAGCATGGGCTGAGATTTATCTGTTAAAGCTTATCTGAT 98519
 QY 1408 AATCTTTGACAAAGTCTTGTGCTACATTTGATGATGATGATGATGATGATGATGAT 1467
 Db 98518 GTTATGTCACACAGTCTTGTGCTACATTTGATGATGATGATGATGATGATGATGAT 98459
 QY 1468 AAGCTTGGATATCTATAGACCCGAAACCGTACACTCCCGATCTTGTGTGTGCGG 1527
 Db 98458 AGGCTGATGCTATTAATTAACCCGAAACCGCATGCTGATCTCCCTGCTGCTGCTGCT 98339
 QY 1528 GCTATGTGATGAGACCTTGTGTCACCTGCTCATCTATGATGATGATGATGATGATGAT 1587
 Db 98338 GTTGTGTCATGAGAACTTGTGCTTCTGATGATGATGATGATGATGATGATGATGAT 98339
 QY 1588 GGTATGACACGGGAACATAGTTCCTCATGTGAGACATTTGTGAGATATCAGCTGGGA 1647
 Db 98338 GGAGCTGATCTGGAAGTGTGCTCCATATGTAAGTATTCGGAAGATATGAGCTGGGG 98279
 QY 1648 CTTAAGCTAGT 1658
 Db 98278 CTTAAGCTAGT 98268

RESULT 13
 LOCUS AP005002 147289 bp DNA linear HTG 28-MAR-2002
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone P0486G03,
 *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AP005002
 VERSION AP005002.1 GI:19773538
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
 clone:P0486G03.
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1

AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
 clone: P0486G03
 JOURNAL Published Only in Database (2002)
 REFERENCE 2 (bases 1 to 147289)
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (27-MAR-2002) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://irp.dna.affrc.go.jp/,
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 Location/Qualifiers
 1. 147289
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="2"
 /clone="P0486G03"

BASE COUNT 39416 a 34573 c 35019 g 38078 t 203 others
 ORIGIN

Query Match 22.5%; Score 373.6; DB 2: Length 147289;
 Best Local Similarity 55.38; Pred. No. 1.7e-84;
 Matches 861; Conservative 0; Mismatches 644; Indels 51; Gaps 5;

QY 149 TAGCTTCCTGCTAGTATGATCTCAATGATCTTTCACCAACACCACTGATTCGA 208
 Db 14100 TAACTTTATCTGTAACCTTATTTGGGTAATGTTCTTTCGAATGACATCTTATCTT 141459
 QY 209 ATCGGATTAAGGGTTTGGCGAAGCTAGAGTTCTCGAGCGGAGTTTCCAAATTCG 268
 Db 14160 CTTGACATATGCAAGATGATGCTGCTGATGATCTAATCTGATCTCAGTACATGCGGTATG 141519
 QY 269 ATAGCTTCGCGAGGCTACTTCTGCTGTTGATGAAATCTTGTGCTTAGAGT 328
 Db 141520 ACAATTTCTGGGGTCTTCTGCAAGAGGATTTGTAAGATCTCTGCGAAGAGGT 141579
 QY 329 ACCAATGATTCATTAACCTTACCTTACCAAGCATCTTCTTATCTCATCTTA 388
 Db 141580 ATGAGTTGACGCTTACCAAGACTTCAAGAAATATCTTCTTATACCTCTTGGAA 141639
 QY 389 AGCTTGAACCTACAGAAAGCTTACAGAGATGAGTGGTGGGTACTGATCTTACAGA 448
 Db 141640 GGTAAAGAGCAAGAAAGCATGTCAGAGAAAGTGTGTCAGGACATAATCATACAG 141699
 QY 449 AAGCTTAACAACTTGATCAAGACA---TATGATGATGATGATGATGATGATGATGAT 505
 Db 141700 AAGCTGGAAGCTGCTGAGTCCAGTGAAGGCTCAACATGACGACAGATGCACTATTC 141759
 QY 506 TTGTTGATTTCTTTAGCGCTTAGGAAACAGATTTCTTACGCTCGCTTTTTC 565
 Db 141760 TGTCTTGAAGATACATGCTGGGTTAGGAAAGATGCTTGAATGCTTCAAGCAATTTTC 141819
 QY 566 TTTAAGCGCTTTTAACGATAGAGCTTGTCTTGAACGAGGAAAGACATGATGATC 625
 Db 141820 TTATAGCACTCTTAAACACCGATTTTGTGACCGCTTATCAGAGATCGGTGACC 141879
 QY 626 TCTTTGCGAGCGCTTCTGATGATGCTGCTGCTTACCTTATGATTTCCATATCAG 685
 Db 141880 TTTCGTGTAACCTTTCACAGAACTGCTGTTGATGCTGGATGCTTCCCTTAA--- 141936
 QY 686 ATCAGTTGATGATTAATCAAGAAATCATCTGTTATGATATATGATGATGATGATGAT 745

Db	141937	---	ACTATGTCGAATTTTACTGAGCTAGGCCAGAGAGCGATATGGAAACATGCTGCAGAAATA	141993			
OY	746	AGGTCATTGATTA	CTAGGGA	CTTTGTC	-----	ATCTTATC	784
Db	141994	AGGTTGTCGGGACAA	TAATACAGATCTGCGCTGTTCTAGGCCCTCCCTACGTTGCC	142053			
OY	785	TTTCATCTGTTTCATGATTA	TGAGATCAATGATAAGATGTTCTTGTCGAAGGAGACAA	844			
Db	142054	TCCTACCTTGATG	TAACTATGATTAATTCATGACACAGCTTTTCTTCTGCGAAGACGACCAAC	142113			
OY	845	CATTTCATCGGGAAG	ATCCCTTGTTGATTGTTTAAACACACAACTTCTGTTCCATCTC	904			
Db	142114	AGTTCTCCAGAGAT	TGCGCGTGTGATCATGAGAACCGACATGTAATCTTATACATACAC	142173			
OY	905	TGTGTTAA	TACCGGTTTCGATGATGAAC	TAAACAGCTATTTCCACAGAAAGCAGCTG	964		
Db	142174	TATTTCTTAATTC	CAAGTTACCAAGATGAGCTCAGAGGCTATTTTCTGCGAAAGATGCTG	142233			
OY	965	TCCTTTCATC	ACTTAAAGTATGATCTTTTTCACACCACTTACCAAGATGAGGCTTAC	1024			
Db	142234	TTTTTTCATC	ACTGTCGACGATATCTTTTCATCCGACAGAAATGCAATGCTATTCAGTTA	142293			
OY	1025	CTAGTACTAC	GACACTTACTTATATCCATTCGCGATGACAGATGGATGGATTCAGTAAG	1084			
Db	142294	AAGGTA	CTACTAGCTTACCTTGC	CAAAAGCTAAATACTGGGAATTCAGATCAGGA	142353		
OY	1085	TTTTTCGATGA	AG-----	ACCGGGCTCACTTTAGATGATGATGATGCACA	1129		
Db	142354	TATTTGAC	AAAGAGGCACTTTGCCAAAAAGGCGATTTTCCATATGCTTTTGAACACA	142413			
OY	1130	TTTTCACTT	GTACTCAAAAGAGAACTTCTAC	TGAA-----GTAGACACACTAGTGG	1183		
Db	142414	TCTCTTCA	TGCGCTCAGATGAAAAAGCTGCTGCCAGAAATCAGTATGAAAGATGAAGCAG	142473			
OY	1184	AGAGATCT	CGGCATGTTATATCCCCCAACACAAAGCCGTGCTGTGCATCTTTGAAG	1243			
Db	142474	AAGCACC	AGCCGCGCCCAAGATTAACACAGACATTTGCTTTTTCACAACTTCTGAGCT	142533			
OY	1244	CGGGTTAC	CGCGAGAACTTAAAGATATGATTTTGGAAATATCCGACATCACTGAGAAA	1303			
Db	142534	CTGTGTAC	AGATGATCAATCCAGAAAGTACAGTAGAGACCCACAGTCATGGAACCA	142593			
OY	1304	TCATCGT	GTTCATACGCCGAGCAAGAGGTTATCAGCAGACCGA	AAAAAAGATGCATA	1363		
Db	142594	GGGTAGA	AGATTTACACGCGCGAGCCATGAGAGTACACAGATCAAAAGAACAGAAAGCACA	142653			
OY	1364	ATGCGA	AAAGCTCTCGGAAATGATCTTTTGAATGACATATATCTTGACAAAGT	1423			
Db	142654	ACATGAA	AGCACTGCGCGAGATCTTCACTTACATGACATGACGACGCTGCTCATCCACAGT	142713			
OY	1424	CTTGCTT	ACTATTTGATATGTAGCTCAAGCTTTTGAAGCTTTAAAGCCTTGATACTCT	1483			
Db	142714	GCTTCTC	ACCTTGGCTACACCGCGAGGCGCTCTCCGGCTCGACCGCTGATCATGT	142773			
OY	1484	ATGAC	CCGAAACCGTACACTCCCGATCTCGTGATGTCGGGCTATGTCGATGAGAC	1543			
Db	142774	TCAGT	CGGAAACACCCCACTGCCGAGCCGCGCTGCGCGCCCACTGTCGATCGAC	142833			
OY	1544	CTTGTTT	CACTCGCTCATTTATGATTTGAAAGCAGAAACGGGTATTGACACGAGGA	1603			
Db	142834	CGTGTCT	CACTACAGGCTCTTCTTCACTGCACTCAAGCGAAGGAAGCAAGCTGAC	142893			
OY	1604	CAC	TAGTTCCTCATGTGACATTTGTGAGATATACGTGGGACTTAAAGCTAGTA	1659			
Db	142894	AGATG	TCGCTGTCGAGGCACTGCGAGGACGTCGATGTTGGGCTGGAAGGTTGTA	142949			

RESULT 14	AC011765/c	LOCUS	DEFINITION
AC011765	134402 bp	DNA	linear
Arabidopsis thaliana chromosome 1 BAC F1M20 genomic sequence, complete sequence.			

ACCESSION	AC011765
VERSION	AC011765.6
KEYWORDS	GI:12324786
SOURCE	HTG.
ORGANISM	Arabidopsis thaliana.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
AUTHORS	1 (bases 1 to 134402) Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Cressy, T.H., Haas, B.J., Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E., Bowman, C.L., White, O., Niernan, W.C. and Fraser, C.M.
TITLE	Arabidopsis thaliana chromosome 1 BAC F1M20 genomic sequence
JOURNL	Unpublished
REFERENCE	2 (bases 1 to 134402) Lin, X. and Kaul, S.
AUTHORS	Direct Submission
TITLE	Submitted (14-OCT-1999) The Institute for Genomic Research, 9712
JOURNL	Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE	3 (bases 1 to 134402) Town, C.D. and Kaul, S.
AUTHORS	Direct Submission
TITLE	Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
JOURNL	Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT	On Jan 19, 2001 this sequence version replaced gi:12280786. Address all correspondence to:ateltigr.org

BAC clone FLM20 is from *Arabidopsis thaliana* chromosome 1. The orientation of the sequence is from SP6 to T7 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GenatxNHM (Mark Borodovsky, http://www.tigr.org/softlab/glimmer_hcm/glimmer.html, and GeneSplicer (Mihaela Perlea and Steven Salzberg, <http://contact.meritallgr.org/>), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/cdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAScan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAScan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smith, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES
source

```

/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F1M20"
complement(1655..3429)
/gene="F1M20.1"
/note="similar to p58 protein kinase GB:AB59449 (Homo sapiens): contains pfam profile: PF00069 Eukaryotic protein kinase domain"
protein_kinase_domain
complement(join<1655..1777,1908..2135,2258..2581,2679..2963,3052..3429))
/gene="F1M20.1"
complement(join(1655..1777,1908..2135,2258..2581,

```



```
Db 25427 GGGAGATCGGCTGACGGATGAAGTTTCGGTTGAGATATGTAATGATGTAATGCCAA 25368
      |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Oy 522 TAGCGGCTTAGGAACAGGACTTCTTCAGCCGCGTTTACCGGCTTTTAC 581
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25367 TGGGATCTTGGGAATGAGATGCTGAGCTGAGCTTCAGCTTCTTTATGCTCTTAAC 25308
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 582 GGATAGAGCTCTGCTGCTTGTGACCGGGAAGACATGATGATCTCTTTCCGAGCCGT 641
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 25307 AATAGAGTTTACTTGTGCAGACTAGAGAGTTGACATGGCTGATCTTTCTCGAGCCATT 25248
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 642 TCTGGATATGCTGCGTGTCTACCTTAGATTTCCCTATGACTGATCATGTTGATGATT 701
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 25247 TCCAAACACTCTGCTGTTCTTCCCGAGAGTTTCCGCTCAACGCCACTTCAACG---- 25192
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 702 AATCAAGAAATCATCTGTTGTTATGATATATGTTGAAGATCAGGATGATTACTGA 761
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 25191 -----AGCAGTCTCTTACGCAATTC 25170
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 762 GGGAACTTGTCTCATCTTATCTTCACTTGTTCATGATGATGAGATCATGATTAAGAT 821
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 25169 TGGCAACCCGATGGTTCGATTCGACATGTAGTTCGTATTCAGTACCAACAAACACT 25110
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 822 GTTCTTCTGTAAGAGACCAACATTCATCGGAAGTCCCTGCTGATTTGTTAAAC 881
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 25109 TTTCTTTGTCAGATAGTCAAGTTTGTGGAGAAACCCCTGTTGATCTTAAAGC 25050
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 882 AGACAATTAATCTTGTTCATCTGCTGTTAATACCGGCTTTCATGATGATGAATCAACAA 941
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 25049 GGAATGTTCTTCTTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 24990
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 942 GCTATTCACAGAAAGACGATGCTTTCATCACTTATGATGATGATCTTTCACCCAAC 1001
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24989 GCTTCTCCGAAAAAAGATATGCTTTCATCTTAACTAGATATCTCTTTCACCCAAC 24930
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1002 TAACCAAGTATGGGCTTACTGATGATGATGATGATGATGATGATGATGATGATGATG 1061
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24929 TAAATGTTGCTGGGCTCATTTACGATGATGATGATGATGATGATGATGATGATGAT 24870
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1062 GAACATTTGGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1121
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24869 AAGATTAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 24810
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1122 GGATCAGATTTCTGATCTGATCTCAAAAAAGAAATCTTACCTGATGATGATGATGATG 1181
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24809 AGATCAGATTTTACTTGTGTAAGTATGATGATGATGATGATGATGATGATGATGATGAT 24750
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1182 GGA---GAGATCTGCGCATGTTAAATACCCCAACAAACGCTGCTGTGATCATCTTT 1238
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24749 AATATCTCCCTCAAGCAAGTTCTGAACCGAAATCAAAAGCAGTGTATCTCATCTTC 24690
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1239 GAACGGGGTTAGCGGGAATTAAGAGATGATGATGATGATGATGATGATGATGATGATG 1298
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24689 TTTCCCGAGGATTTTAAAGATTAATGATGATGATGATGATGATGATGATGATGATGATG 24630
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1299 AGAATCATCGGTTTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1358
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24629 AGAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 24570
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1359 GCATTAATGCGCAAGCTCTGCGGAATGATCTTTGAGTTGACAGATTAATCTTTG-- 1416
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24569 GGAAGAGTAAGAGCATGCGCTGATATATACCTTCTGAGTTGTTGATGCGGTGTTGTT 24510
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1417 -ACAAGTCTGCTCATATTTGATATGATGATGATGATGATGATGATGATGATGATGATG 1475
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24509 CACAGGTTTATGCTCTACCTGCTGAGGATGCTCATGCGCTTGAAGGTTGAAGCCATG 24450
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1476 GATACCTATAGACCCGAAACCTTACAACTCCGATCTCTGCTGCTGCTGCTGCTGCTGCT 1535
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 24449 GGTGTGAACAACACTGGAATGGGATGCGCCATGAGCTTACTGTGGAAGCAAGATC 24390
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Oy 1536 GATGAGCTTGTTCACACTGCGCTCATTTGATGATGATGATGATGATGATGATGATGATG 1579
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
```

```
Db 24389 AATAGAGCCCTGTTCCCAAGCAGATTTGTCATGCTGTAAG 24346
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
RESULT 15
AF417473 1706 bp mRNA linear PLN 27-DEC-2001
LOCUS Arabidopsis thaliana fucosyltransferase-1-like protein FUT3 (FUT3)
DEFINITION
ACCESSION AF417473
VERSION AF417473.1 GI:17980418
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana.
Arabidopsis thaliana.
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
Rosidae: eucosids II: Brassicales: Brassicaceae: Arabidopsis.
1 (bases 1 to 1706)
Saaria, R., Wagner, T.A., O'Neill, M.A., Paik, A., Wilkerson, C.G.,
Keegstra, K. and Raikhel, N.V.
Characterization of a family of Arabidopsis genes related to
xyloglucan fucosyltransferase1
Plant Physiol. 127 (4), 1595-1606 (2001)
JOURNAL MEDLINE 21608393
REFERENCE PUBMED 11743104
AUTHORS Saaria, R., Keegstra, K. and Raikhel, N.V.
TITLE Submitted (07-SEP-2001) PRL, Michigan State University, 122 Plant
Biology Building, East Lansing, MI 48824, USA
FEATURES
source
1. 1706
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/map="At1g74420"
1. 1706
/gene="FUT3"
62. 1543
/gene="FUT3"
/note="AtFUT3"
/codon_start=1
/product="fucosyltransferase-1-like protein FUT3"
/db_xref="GI:17980418"
/translation="MMKPSLSMKTMGLLAVLVAVMSLSVLLDPSPDIYE
AASRVLOSRLHDDGUSEKRAQJENINLVPSFDESCLSRKESLVKRESFQSYL
DYRLQRYEDLHRRCGPPTRSYNLTLDLKSGRDSEVSGCRYVILNMSGDLNRL
SLASAPLYALITNRLFLVELGVADLFCFEPMTWTFPLPEFLNHSFMSQILRS
GNMVAFLRVHVRDSDQKLFECDSQVLEETFWLILKADSFLPSLFSVSPKOL
OMLEPEEDTAEHLISQYLFHPNTNVMGLITRYVAVLAKDORIGIIGVSESGNOF
OHLDIOLACGTRHKLLEPYDKORNLPSQVILNRSKAVFISSSPGFYSIRPYVE
NPYVMGIIISYHRSYIDYOKTTPNMSKRMATITLSCSDALVVTGLMSLVEVAH
GLGGLKRWLVLMKAENGTAHEPYCVKANSIEPCSOATLPHCKKD"
BASE COUNT 476 a 345 c 387 g 498 t
ORIGIN
Query Match 20.7% Score 344.2 DB 8 Length 1706;
Best local similarity 56.3% Pred. No. 66-77;
Matches 761; Conservative 0; Mismatches 533; Indels 57; Gaps 4;
```


GenCore version 5.1.4-g5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 09:59:34 ; Search time 383 Seconds
(without alignments)
9772.379 Million cell updates/sec

Title: US-10-037-311A-2
Perfect score: 1662
Sequence: 1 atggatcagatcgtacag.....gggacttaagctatgta 1662

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	604.8	36.4	1720	21 AAC50203	Arabidopsis thaliana
2	248.4	14.9	875	21 AAC34268	Arabidopsis thaliana
3	50.6	3.0	171	24 ABL76082	Corn tassell-derivate
4	43.4	2.6	7001	24 ABL33919	Human DNA for strag
5	40.6	2.4	6430	24 ABLN80122	Human chemically m
6	40	2.4	32145	22 AAK68491	Human immune/haema
7	40	2.4	32145	22 AAK68575	Human immune/haema
8	39.4	2.4	6397	24 ABL32819	Human immune syste
9	38.8	2.3	4590	22 AAL24065	Yeast AOD9604-asso

10	38.2	2.3	50000	24	ABL56202	AMEPV genome fragm
11	37.8	2.3	1962	24	AAS62336	cDNA sequence #123
12	37.8	2.3	2839	22	AAF58335	Human GRP-blinding
13	37.8	2.3	9516	22	AAS42103	Genomic sequence #
14	37.6	2.3	6392	24	ABL32685	Human immune syste
15	37.6	2.3	6392	24	ABL34507	Human metastasis a
16	37	2.2	792	21	ABLN81133	Shrimp polynucleot
17	37	2.2	15225	19	AAV17552	Respiratory syncyt
18	37	2.2	15225	21	AAB88744	Respiratory syncyt
19	37	2.2	15225	21	AAD25520	Human respiratory
20	37	2.2	17967	24	ABL33014	Human immune syste
21	36.8	2.2	15229	19	AAV18276	RSV isolate 18537
22	36.8	2.2	15229	20	AAZ22910	Nucleotide sequenc
23	36.8	2.2	15229	20	AAX35268	DNA encoding the l
24	36.6	2.2	780	21	AAC36958	Arabidopsis thalia
25	36.6	2.2	1022	24	ABN98442	Arabidopsis thalia
26	36.6	2.2	1198	22	AAK82621	Human immune/haema
27	36.6	2.2	6277	22	AAS46324	Tumour suppressor
28	36.2	2.2	5271	21	AAA28051	Human androgen shu
29	36.2	2.2	5355	21	AAA28052	Human androgen shu
30	36.2	2.2	23683	24	ABL70482	Chemically treated
31	36.2	2.2	23683	24	ABL34623	Human metastasis a
32	36	2.2	639	24	ABLQ6046	Arabidopsis thalia
33	36	2.2	687	21	AAC37908	Arabidopsis thalia
34	36	2.2	688	21	AAC40855	Arabidopsis thalia
35	36	2.2	6060	18	AAV74845	Staphylococcus aur
36	36	2.2	7025	24	ABK4059	Human chemically p
37	36	2.2	7025	24	AAS63350	Chemically pretrea
38	36	2.2	9789	17	AAT41852	CDNA encoding Plas
39	36	2.2	10328	24	ABL33544	Human immune syste
40	35.8	2.2	683	22	AAK80819	Human immune/haema
41	35.6	2.1	540	23	ABLV5465	Human prostate exp
42	35.6	2.1	6874	22	AAS45440	Chemically pretrea
43	35.6	2.1	6874	24	ABN80230	Human chemically m
44	35.6	2.1	6874	24	ABL70381	Chemically treated
45	35.6	2.1	6874	24	AAS61333	Human gene regulat

ALIGNMENTS

RESULT 1	
AAC50203	
ID AAC50203 standard; DNA; 1720 BP.	
AC AAC50203:	
XX 18-OCT-2000 (first entry)	
DT Arabidopsis thaliana DNA fragment SEQ ID NO: 63957.	
XX DE	
XX Arabidopsis thaliana	
XX DE	
XX Hybridisation assay; genetic mapping; gene expression control;	
KW protein identification; signal transduction pathway;	
KW metabolic pathway; promoter; termination sequence; ss.	
XX Arabidopsis thaliana.	
OS Arabidopsis thaliana.	
PN EP1033405-A2.	
XX 06-SEP-2000.	
PD 25-FEB-2000; 2000EP-0301439.	
PF 25-FEB-1999;	99US-0121825.
XX 05-MAR-1999;	99US-0123180.
PR 09-MAR-1999;	99US-0123548.
PR 23-MAR-1999;	99US-0125788.
PR 25-MAR-1999;	99US-0126264.
PR 29-MAR-1999;	99US-0126785.
PR 01-APR-1999;	99US-0127462.
PR 06-APR-1999;	99US-0128234.
PR 08-APR-1999;	99US-0128714.
PR 16-APR-1999;	99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130044.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136382.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138647.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147419.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161982.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 36.4%; Score 604.8; DB 21; Length 1720;

Best Local Similarity 66.0%; Pred. No. 4.6e-168; Matches 946; Conservative 0; Mismatches 467; Indels 21; Gaps 4;

OY 232 GCTAGAGTTCGACGCGGAGTTCCTCCAAATTCGTATACCTTCGAGGCGCTACTT 291
DB 263 GCTAATCAACAGCGATCCGAACACCGGGCGGATAGACTAATAGAGGCGCTTTTA 322
OY 292 GCTCTGCTTTGAGAGATCTTCGCTTACTAGTAC---AATCAGTTATTAACGT 348
DB 323 ACCGAGATTCGATGAGAGGTTCTTGCTGATAGTATCATATAAATCTTCTTGATCGC 382
OY 349 AAACCTTCACCTTACAGCCATCTTCTATCTCATCTTAAGCTTAGAAGTACGAAAG 408
DB 383 AAGCTTCACCATACAAACCGCTGAATATCTGTCTCGAAGCTTAGAAGTATGAGATG 442
OY 409 CTTCACAAGCGATGTGTCCGGGTACTGATCTTACAAAGAAAGCTCTAAACACTTGAT 468
DB 443 CTTCACAAGCGATGTGTCCGGGTACTGATCTTACAAAGAAAGCTCTAAACACTTGAT 502
OY 469 CAGAACAATATGATGAT-----GATGATGATGATGATGATGATGATGAT 516
DB 503 CATGATGAGATTAATTAATGCAAGCAAAATCAGATGATGATGATGATGATGATGAT 562
OY 517 TCTTTAAGCGGCTTAGGAGACAGATACTTCTTACGCTCGGTTTCTTTACCGGCTT 576
DB 553 GCTGATTCAGGCGCTTAGGAGACCGACTACTCTCTGCTCTGCTCTCTACGCTCTG 622
OY 577 TTACGATAGAGTCTTCTGTTGACCGAGGAGAAAGACATGATGATCTTTTTCGAG 636
DB 623 TTGACTGATGATGATCTTCTGTTGACCAACCGCAAGATGATGATGATCTTATGCGAG 682
OY 637 CCGTTTCTGCGATGATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 696
DB 663 CCATTTTCAGGCTTACGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 742
OY 697 GGAATTAATTAAGATCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 756
DB 743 GGAATTAATTAAGATCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 802
OY 757 ACTGAGGAGACTTCTGCTATCTTATCTTCACTCTTCTTCACTGATGATGATGATGAT 816
DB 803 TCGACTTATCTTCCCGCCACATCTATATGATTAACCTTATGATTAAGGATGATGAT 862
OY 817 AAGATGTTCTTCTGAGAGAGACCAACATTCATCGGAGAGTCCCTTGGTGTATGTT 876
DB 863 AAGATGTTCTTCTGAGAGAGATCAAGGTTGATGATGATGATGATGATGATGATGAT 922

OY 877 AAACAGACAAATTAATCTTGTTCATCTCTGTGTTAAATACCGGGTTTCGATGATGACTA 936
DB 923 AAGAGCAATGTTTACTTGTTCATCTTATTAATCAACATTTCCAAACCGAGACTA 982
OY 937 AACAACTATTTCCACAGAGAGGAGCTGTCTTCACTACTAGTATGATGATGATGATGAT 996
DB 983 ACGAAGCTGTTTCCCGCAGAAAGAACCGGTTTCACTACTGTTGCTGATCTTCTTTCAC 1042
OY 997 CCACTAACCAAGTATGAGGCTTACGATGATGATGATGATGATGATGATGATGATGATG 1056
DB 1043 CCTAAATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1102
OY 1057 GATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1116
DB 1103 GATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1162
OY 1117 GATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1176
DB 1163 GATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1222
OY 1177 CTAGTGAAGATCTTCCGATGT---TAATACCCCAACAAAGCCGTGCTTGCACA 1233
DB 1223 CAAGAAGATCAAAAGTCAACATATCGAACATTCACAAAAGTAAAGCTTCTTGTCACT 1282
OY 1234 TCTTGAACGCGGTTACCGGAGAACTTAAGAGATGATGATGATGATGATGATGATGATG 1293
DB 1283 TCTTGTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1342
OY 1294 ACTGAGAAATCATCGGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1353
DB 1343 ACGGAGATGATTAATTAAGTATATGATGATGATGATGATGATGATGATGATGATGAT 1402
OY 1354 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1413
DB 1403 AAGTTCACGACCAAAAGGCGCTGCTGATGATGATGATGATGATGATGATGATGATG 1462
OY 1414 GTGACAAAGTCTTGTGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1473
DB 1463 GTTGCAGATCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1522
OY 1474 TGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1533
DB 1523 TGGTATCTTATCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1582
OY 1534 TCGATGAGAGCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1590
DB 1583 TCGATGAGAGCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1642
OY 1591 ATTGACACGGAAGCACTAGTCTCTCATGAGACATTTGAGAGATPTCAGCTGG 1644
DB 1643 ACTGATGCGGGAAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1696

RESULT 2
AAC34268
ID AAC34268 standard; DNA; 875 BP.

AC AAC34268;
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 6044.
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 6044.
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
XX
XX 06-SEP-2000.
XX

PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

XX S0 Sequence 171 BP; 41 A; 52 C; 53 G; 24 T; 1 other;
 Query Match 3.0%; Score 50.6; DB 24; Length 171;
 Best Local Similarity 55.9%; Pred. No. 0.0001;
 Matches 95; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 QY 1219 GCGGTGCTTGTACATCTTGAACGCGGTTACCGGAGAACTTAAGATGATGTATGG 1278
 1 GCGGTGCTTGAATCGGCTCTACTGCGGTACACAGAAAGATCCCGCATGTACTAC 60
 Db 1279 GAATATCCGACATCACTGAGAAATATATGCGTTCATCAGCCGACCAAGAAGTTAT 1338
 61 GAGANCCCAACCAAGACTGGGAGATCGTGGTTCAGCCCAAGCCAGAGACAG 120
 QY 1339 CAGCAGACCAAAAAAAGATGCAATATGCAAGCTCTTCGGGAATGTA 1388
 121 CAGCAGTACACGTCACCAACGACCAACCAAGAGCGCTGGCCGAGATCTA 170
 Db
 RESULT 4
 ABR33919
 ID ABR33919 standard; DNA; 7001 BP.
 XX ABR33919;
 AC
 DT 18-JUN-2002 (first entry)
 XX
 DE Human DNA for staging of Astrocytomas #1.
 XX
 KM Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; Cpg;
 KW bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;
 XX matrix assisted laser desorption/ionization mass spectrometry.
 OS Homo sapiens.
 XX
 PN WO200202808-A2.
 PD 10-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07538.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI: 2002-171649/22.
 XX
 PT Novel chemically modified genomic DNA sequences, useful in the
 PT characterisation, classification, differentiation, grading, staging,
 PT treatment and/or diagnosis of astrocytomas or predisposition to
 PT astrocytomas
 PS
 Claim 1; SEQ ID No 1; 37pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence (I) of at
 CC least 18 bases in length of a segment of chemically pre-treated genomic
 CC DNA which has any one of the sequences of (ABK33919-ABK34032) or its
 CC complement. Also included are an oligonucleotide or peptide nucleic
 CC acid (or set thereof) of at least 9 nucleotides which hybridises to (I),
 CC primers for (I), probes for detecting cytosine methylation or single-
 CC nucleotide polymorphisms (SNP) in (I), an array of oligomers
 CC or peptide nucleic acids for analysing diseases associated with the
 CC methylation states of the Cpg dinucleotides of (I). The array is useful
 CC for determining genetic and/or epigenetic parameters, classification,
 CC differentiation, grading, staging, treatment and/or diagnosis of
 CC astrocytomas, or the predisposition to astrocytomas by analysing cytosine
 CC methylations, involves obtaining a biological sample containing genomic
 CC DNA, extracting the genomic DNA, converting cytosine bases which are
 CC unmethylated at the 5-position, in the genomic DNA sample, to uracil or

CC another base which is dissimilar to cytosine in terms of hybridisation
 CC behaviour, by chemical treatment and amplifying chemically pre-treated
 CC genomic DNA fragments using the array and a polymerase, where the
 CC amplificates carry a detectable label. The method further involves
 CC identifying methylation status of one or more cytosine positions, and
 CC analysing methylation status of the cytosine positions by reference to
 CC one or more data sets. The genomic DNA is chemically treated by using a
 CC bisulphite, hydrogen sulphite or disulphite. The amplification
 CC step amplifies DNA which is of particular interest in astrocytoma or
 CC brain tissue, based on the specific genomic methylation status of brain
 CC tissues, as opposed to background DNA. The amplificates carry a
 CC fluorescent label or radionuclide. Optionally the labels of the
 CC amplificates are detachable molecule fragments having a typical mass
 CC which are detected in a mass spectrometer. The fragments of chemically
 CC pre-treated genomic DNA to be amplified, have a single positive or
 CC negative charge for a better detectability in the mass spectrometer.
 CC Preferably, the amplificates or fragments of the amplificates are
 CC detected by matrix assisted laser desorption/ionization mass spectrometry
 CC (MALDI) or using electron spray mass spectrometry (ESI). The
 CC present sequence is one of the chemically pre-treated reference DNA
 CC samples of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcl_sequences.
 XX
 S0 Sequence 7001 BP; 2309 A; 40 C; 1249 G; 3403 T; 0 other;
 Query Match 2.6%; Score 43.4; DB 24; Length 7001;
 Best Local Similarity 54.8%; Pred. No. 0.081;
 Matches 86; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
 QY 647 GATATCGTGGTGTGCTACTTACCTTCCCTATGACGATGATGATGATTAATC 706
 2610 GTGTGTGTGTATTTTGTGATATTTAAATTTGTATAGAGATGTTTACGTTAATTT 2669
 Db
 QY 707 AAGATCATCTGCTGTATGATATATGATGCAAGATCAGGTATGATGATGAGGAA 766
 2670 ATGAGTATTTAGTTTGTGGGTTTATGATTTGATTTATTTGGAAATGATATTTTACG 2729
 Db
 QY 767 CTTTGTCTCATCTTATCTTCATCTCTGTTCATGATTA 803
 2730 GTTGTATTAATTAATTAATTTTGTGTTAAATTTA 2766
 Db
 RESULT 5
 ABR80122
 ID ABR80122 standard; DNA; 6430 BP.
 XX
 AC ABR80122;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human chemically modified disease associated gene SEQ ID NO 139.
 XX
 KM Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
 KW antidiabetic; cytostatic; anticonvulsant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200200927-A2.
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07536.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI: 2002-130908/17.
XX
XX Novel nucleic acid useful for diagnosis and therapy of diseases
XX associated with development genes such as diabetes, comprises a
XX sequence of a segment of chemically pretreated DNA of genes associated
XX with development
XX
XX Claim 1: SEQ ID NO 139; 27pp; English.
XX
XX The invention relates to a nucleic acid (I) comprising a sequence at
XX least 18 bases in length of a segment of chemically pretreated DNA (II)
XX of genes associated with development selected from 87 genes listed in
XX the specification such as ACCFN, ADFN, or AFDI and comprising one of 350
XX sequences (ABN79984-ABN80333) or their complements. The invention is
XX useful for the diagnosis or therapy of diseases associated with
XX development genes, in particular disease related to homeobox containing
XX genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
XX associated with congenital heart disease, epilepsy, diseases related to
XX histone deacetylation, Currarino syndrome, diseases related with the
XX development of the brain and limb girdle muscular dystrophy and dwarfism.
XX Oligomers specific to each of the genes are useful for detecting the
XX methylation state of all CpG dinucleotides within the 350 sequences or
XX (II) and their complementary sequences, as primer oligonucleotides for
XX the amplification of the 350 sequences, (II) and/or their complements and
XX as oligomer probes for detecting the cytosine methylation state and/or
XX single nucleotide polymorphisms (SNPs).
XX Note: The sequence data for this patent did not form part of the printed
XX specification but is based on sequence information supplied to Derwent by
XX the European Patent Office.
XX
XX Sequence 6430 BP; 1334 A; 334 C; 1866 G; 2896 T; 0 other:
SQ
Query Match 2.4%; Score 40.6; DB 24; Length 6430;
Best Local Similarity 45.5%; Pred. No. 0.52;
Matches 145; Conservative 0; Mismatches 174; Indels 0; Gaps 0;
QY 561 TTTTCTTAAAGCGCTTAAAGGATAGAGTCTGCTGTGACCGGAGGAAGACATGGA 620
DB 1583 TTAAGTTTAAATAAAGTTGATCGATTGATTTTAACTGTTTAACTGTTAAATGGGGA 1642
QY 621 TGATCTCTTTTCGAGCCGTTCTCGTATGCGTGTGATGCTTACCTTTTGATTTCCCTAT 680
DB 1643 TTGGATTGTTGGTATGTTTTCGCGTGTGTTTTCCTTTTTCCTTTTATATATATAT 1702
QY 681 GACTGATCAGTTTGATGATTAATCAAGAAATCATCTGCTTATGATATATGCTGAA 740
DB 1703 AATATTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTT 1762
QY 741 GAATCAGGATGATGATGAGGAGGACTTGTCTCATCTTTATCTTCATCTGTTTCATGA 800
DB 1763 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1822
QY 801 TTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 860
DB 1823 TCGTATATATATTTTAAACGTTTGAATGTAAGAGGTTTAACTTTTTCGAGCTTA 1882
QY 861 CCGTTGGTTGATGTTAA 879
DB 1883 ATCGTGGTTTGTATTTAA 1901
RESULT 6
AAK68491/C
ID AAK68491 standard; DNA; 32145 BP.
XX AAK68491;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23303.

XX
XX Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0228287.
XX 01-SEP-2000; 2000US-0228343.
XX 01-SEP-2000; 2000US-0228344.
XX 01-SEP-2000; 2000US-0228345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236882.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
PR 05-JUN-2001; 2001US-0239678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM:
XX WPI; 2001-483426/52.
DR
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
PS Disclosure; SEQ ID NO 23303; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (II)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 other;
Query Match 2.4%; Score 40; DB 22; Length 32145;
Best Local Similarity 47.9%; Pred. No. 1.7;
Matches 115; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 692 TTGATGATTAATCAATCAGATCATCTGTTATGATATATGTTGAGATCAGTGA 751
DB 11348 TTTAGGATATTTTTCAGCAGTTTCTATGATGATGATGATTTTCCAACTTTGG 11289
QY 752 TTGATACGAGGAGACTTGTCTCATCTTATCATCTCATCTTATATATGATGAGATC 811
DB 11288 CAAATTAGCAATCTTTTATATACATGATGATTTGCTTTTGTATGAGGAGGAAAC 11229
QY 812 ATGATPAGATGTTCTCTGTGAAGGAGACCAACATTCATCGGGAAGTCCCTTGTTGA 871
DB 11228 TTGTGAAATCAACATTTGGAATTAAGAGTTATAGGCTCATTTAGTAATGCGCATTAATTA 11169
QY 872 TTGTTAAACAGACAAATACCTTTGTCATCTCTGTGTATATACCGGTTTCGATGATG 931
DB 11168 TTAATTAGGCTAAACACGTAATTTCTCTTTTGTGATGATGAGATTTCCCTCTTG 11109
RESULT 7
AAK68575/c
ID AAK68575 standard; DNA; 32145 BP.
XX
XX AAK68575;
AC
AC
DT 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23387.
DE Human immune/haematopoietic; immune/haematopoietic antigen; cancer;
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX
XX W0200157182-A2.
XX
XX

BD	09-AUG-2001.	2001WO-USO1354.
XX	17-JAN-2001.	2001WO-USO1354.
XX	31-JAN-2000.	2000US-0179065.
PR	04-FEB-2000.	2000US-0180628.
PR	24-FEB-2000.	2000US-0184664.
PR	02-MAR-2000.	2000US-0186350.
PR	16-MAR-2000.	2000US-0190874.
PR	17-MAR-2000.	2000US-0190976.
PR	18-APR-2000.	2000US-0198123.
PR	19-MAY-2000.	2000US-0205051.
PR	07-JUN-2000.	2000US-0209467.
PR	28-JUN-2000.	2000US-0211886.
PR	30-JUN-2000.	2000US-0215136.
PR	07-JUL-2000.	2000US-0216647.
PR	07-JUL-2000.	2000US-0216880.
PR	11-JUL-2000.	2000US-0217487.
PR	11-JUL-2000.	2000US-0217496.
PR	14-JUL-2000.	2000US-0218290.
PR	26-JUL-2000.	2000US-0220964.
PR	26-JUL-2000.	2000US-0220964.
PR	14-AUG-2000.	2000US-0224518.
PR	14-AUG-2000.	2000US-0224519.
PR	14-AUG-2000.	2000US-0225213.
PR	14-AUG-2000.	2000US-0225214.
PR	14-AUG-2000.	2000US-0225266.
PR	14-AUG-2000.	2000US-0225267.
PR	14-AUG-2000.	2000US-0225268.
PR	14-AUG-2000.	2000US-0225270.
PR	14-AUG-2000.	2000US-0225447.
PR	14-AUG-2000.	2000US-0225457.
PR	14-AUG-2000.	2000US-0225758.
PR	14-AUG-2000.	2000US-0225758.
PR	18-AUG-2000.	2000US-0226279.
PR	22-AUG-2000.	2000US-0226681.
PR	22-AUG-2000.	2000US-0226681.
PR	23-AUG-2000.	2000US-0227182.
PR	30-AUG-2000.	2000US-0227109.
PR	30-AUG-2000.	2000US-0228924.
PR	01-SEP-2000.	2000US-0228287.
PR	01-SEP-2000.	2000US-0229343.
PR	01-SEP-2000.	2000US-0229344.
PR	01-SEP-2000.	2000US-0229344.
PR	05-SEP-2000.	2000US-0229509.
PR	05-SEP-2000.	2000US-0229513.
PR	06-SEP-2000.	2000US-0230437.
PR	06-SEP-2000.	2000US-0230438.
PR	08-SEP-2000.	2000US-0231242.
PR	08-SEP-2000.	2000US-0231243.
PR	08-SEP-2000.	2000US-0231244.
PR	08-SEP-2000.	2000US-0231413.
PR	08-SEP-2000.	2000US-0231414.
PR	08-SEP-2000.	2000US-0232080.
PR	12-SEP-2000.	2000US-0233081.
PR	12-SEP-2000.	2000US-0233168.
PR	14-SEP-2000.	2000US-0233397.
PR	14-SEP-2000.	2000US-0233398.
PR	14-SEP-2000.	2000US-0233399.
PR	14-SEP-2000.	2000US-0233423.
PR	14-SEP-2000.	2000US-0233424.
PR	14-SEP-2000.	2000US-0233440.
PR	14-SEP-2000.	2000US-0233463.
PR	14-SEP-2000.	2000US-0233064.
PR	14-SEP-2000.	2000US-0233065.
PR	14-SEP-2000.	2000US-0233065.
PR	21-SEP-2000.	2000US-0234223.
PR	21-SEP-2000.	2000US-0234274.
PR	25-SEP-2000.	2000US-0234997.
PR	25-SEP-2000.	2000US-0234998.
PR	26-SEP-2000.	2000US-0235484.
PR	27-SEP-2000.	2000US-0235834.
PR	27-SEP-2000.	2000US-0235836.
PR	29-SEP-2000.	2000US-0236327.
PR	29-SEP-2000.	2000US-0236327.

PR	29-SEP-2000;	2000US-0236368	
PR	29-SEP-2000;	2000US-0236369	
PR	29-SEP-2000;	2000US-0236370	
PR	20-OCT-2000;	2000US-0236802	
PR	20-OCT-2000;	2000US-0237037	
PR	02-OCT-2000;	2000US-0237038	
PR	02-OCT-2000;	2000US-0237039	
PR	02-OCT-2000;	2000US-0237040	
PR	13-OCT-2000;	2000US-0239935	
PR	13-OCT-2000;	2000US-0239937	
PR	20-OCT-2000;	2000US-0240960	
PR	20-OCT-2000;	2000US-0241121	
PR	20-OCT-2000;	2000US-0241185	
PR	20-OCT-2000;	2000US-0241786	
PR	20-OCT-2000;	2000US-0241787	
PR	20-OCT-2000;	2000US-0241809	
PR	20-OCT-2000;	2000US-0241809	
PR	01-NOV-2000;	2000US-0244617	
PR	08-NOV-2000;	2000US-0246474	
PR	08-NOV-2000;	2000US-0246475	
PR	08-NOV-2000;	2000US-0246476	
PR	08-NOV-2000;	2000US-0246477	
PR	08-NOV-2000;	2000US-0246478	
PR	08-NOV-2000;	2000US-0246523	
PR	08-NOV-2000;	2000US-0246524	
PR	08-NOV-2000;	2000US-0246525	
PR	08-NOV-2000;	2000US-0246526	
PR	08-NOV-2000;	2000US-0246527	
PR	08-NOV-2000;	2000US-0246528	
PR	08-NOV-2000;	2000US-0246532	
PR	08-NOV-2000;	2000US-0246609	
PR	08-NOV-2000;	2000US-0246610	
PR	08-NOV-2000;	2000US-0246611	
PR	08-NOV-2000;	2000US-0246613	
PR	17-NOV-2000;	2000US-0249207	
PR	17-NOV-2000;	2000US-0249208	
PR	17-NOV-2000;	2000US-0249209	
PR	17-NOV-2000;	2000US-0249210	
PR	17-NOV-2000;	2000US-0249211	
PR	17-NOV-2000;	2000US-0249212	
PR	17-NOV-2000;	2000US-0249213	
PR	17-NOV-2000;	2000US-0249214	
PR	17-NOV-2000;	2000US-0249215	
PR	17-NOV-2000;	2000US-0249216	
PR	17-NOV-2000;	2000US-0249217	
PR	17-NOV-2000;	2000US-0249218	
PR	17-NOV-2000;	2000US-0249219	
PR	17-NOV-2000;	2000US-0249220	
PR	17-NOV-2000;	2000US-0249221	
PR	17-NOV-2000;	2000US-0249222	
PR	17-NOV-2000;	2000US-0249223	
PR	17-NOV-2000;	2000US-0249224	
PR	17-NOV-2000;	2000US-0249225	
PR	17-NOV-2000;	2000US-0249226	
PR	17-NOV-2000;	2000US-0249227	
PR	17-NOV-2000;	2000US-0249228	
PR	17-NOV-2000;	2000US-0249229	
PR	17-NOV-2000;	2000US-0249230	
PR	01-DEC-2000;	2000US-0250160	
PR	01-DEC-2000;	2000US-0250381	
PR	05-DEC-2000;	2000US-0251030	
PR	05-DEC-2000;	2000US-0251030	
PR	05-DEC-2000;	2000US-0251988	
PR	06-DEC-2000;	2000US-0256719	
PR	06-DEC-2000;	2000US-0251479	
PR	06-DEC-2000;	2000US-0251856	
PR	08-DEC-2000;	2000US-0251868	
PR	08-DEC-2000;	2000US-0251869	
PR	08-DEC-2000;	2000US-0251989	
PR	08-DEC-2000;	2000US-0251990	
PR	11-DEC-2000;	2000US-0254097	
PR	05-JAN-2001;	2001US-0259678	
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Rosen CA, Barash SC, Ruben SM,		
XX	WPI: 2001-484362/52.		

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM,
WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 PS Disclosure; SEQ ID NO 23387; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC treatment of diseases associated with inappropriate (I) expression and
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 32145 BP; 8777 A; 7496 C; 7117 G; 8755 T; 0 other;
 Query Match 2.4%; Score 40; DB 22; Length 32145;
 Best Local Similarity 47.9%; Pred. No. 1.7;
 Matches 115; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
 QY 692 TTGATGATTAATCAAGATCATCTGCTGTTATGATGATATGATGAGATCAGTGA 751
 DB 11348 TTAAAGGATATTTTATGACCACTTTTATGATGATGATATGATGATGATGATG 11289
 QY 752 TTGATGATGAGGAGACCTGCTCATCTTATCTCATCTGTTCAATGATGAGATC 811
 DB 11288 CAATATGAAATCTTTTATATACATGATGATGATGATGATGATGATGATGATG 11229
 QY 812 ATGATAGATGTTCTCTGTGAAGACCAACATTCATCGGAAAGTCCCTGGTGA 871
 DB 11228 TTGTAGAAATCAACATGGAATAGATTAAGCTCATTAAGTAATGCGCATTAATTA 11169
 QY 872 TTGTTAAACAGACATTAATCTGTCATCTCTGTGTTAATACCGGCTTCGATGATG 931
 DB 11168 TTATATGAGCTAAAACTGCTAATTCCTCTTTTGTGAGATGAGATTCGCTCTTG 11109
 RESULT 8
 ABL32819
 ID ABL32819 standard; DNA; 6397 BP.
 AC ABL32819;
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 792.
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antileukemic; antineoplastic; antineoplastic; antineoplastic;
 KW antineoplastic; anti-HIV; anticonvulsant; ophthalmological;
 KW antineoplastic; antineoplastic; antidiabetic; antidiabetic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 OS
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX

PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPig-) EPiGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 XX
 CC Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 PS Claim 1; SEQ ID NO 792; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 6397 BP; 1749 A; 73 C; 1302 G; 3273 T; 0 other;
 Query Match 2.4%; Score 39.4; DB 24; Length 6397;
 Best Local Similarity 57.9%; Pred. No. 1.2;
 Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
 QY 559 GTTTTCTTTACGCGCTTTAAGCATAGAGTCTGCTTTGACCGAGGAAACATG 618
 DB 1141 GTTTTCTTTACGCGCTTTTGTAGAGTCTGCTTTGACCGATGAATGAGATG 1200
 QY 619 GATGATCTCTTTGCGAGCGCTTCTCGTATGCTGCTGCTGCTGCTGCTGCTGCTG 678
 DB 1201 GATGATCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 QY 679 A 679
 DB 1261 A 1261
 RESULT 9
 AAH24065/C
 ID AAH24065 standard; DNA; 4590 BP.
 AC AAH24065;
 DT 29-AUG-2001 (first entry)
 DE Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
 XX Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
 KW modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
 KW lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
 KW functional food; transgenic yeast; fat/lean ratio; food use; ds.
 XX Saccharomyces cerevisiae.
 OS
 XX
 XX Key Location/Qualifiers
 FT misc_feature 10
 FT /*tag= a
 FT /note= "Represented as * in the specification"
 FT 3617
 FT /*tag= b
 FT /note= "Represented as * in the specification"
 FT misc_feature 3649
 FT /*tag= c
 FT /note= "Represented as * in the specification"
 FT misc_feature 3679

FT	misc_feature	/*tag= d	"Represented as * in the specification"
FT	misc_feature	/note= 3819	
FT	misc_feature	/*tag= e	"Represented as * in the specification"
FT	misc_feature	/note= 3862	
FT	misc_feature	/*tag= f	"Represented as * in the specification"
FT	misc_feature	/note= 3864	
FT	misc_feature	/*tag= g	"Represented as * in the specification"
FT	misc_feature	/note= 3888	
FT	misc_feature	/*tag= h	"Represented as * in the specification"
FT	misc_feature	/note= 3890	
FT	misc_feature	/*tag= i	"Represented as * in the specification"
FT	misc_feature	/note= 3912	
FT	misc_feature	/*tag= j	"Represented as * in the specification"
FT	misc_feature	/note= 3914	
FT	misc_feature	/*tag= k	"Represented as * in the specification"
FT	misc_feature	/note= 3938	
FT	misc_feature	/*tag= l	"Represented as * in the specification"
FT	misc_feature	/note= 3939	
FT	misc_feature	/*tag= m	"Represented as * in the specification"
FT	misc_feature	/note= 3941	
FT	misc_feature	/*tag= o	"Represented as * in the specification"
FT	misc_feature	/note= 3943	
FT	misc_feature	/*tag= p	"Represented as * in the specification"
FT	misc_feature	/note= 4361	
FT	misc_feature	/*tag= q	"Represented as * in the specification"
PN	W0200133977-A1.		
PD	17-MAY-2001.		
PF	06-NOV-2000; 2000WO-AU01362.		
PR	05-NOV-1999; 99AU-0003875.		
PA	(META-) METABOLIC PHARM LTD.		
PI	Belyea CI, Ng FM, Vaughan P;		
PI	WPI; 2001-328676/34.		
PS	New organisms containing nucleic acid encoding a growth hormone fragment which modulates lipid metabolism are useful to produce dietary aids for obesity and in the meat production industry -		
XX	Disclosure; Page 48-50; 54pp; English.		
XX	The invention relates to novel transgenic organisms useful in the production of functional food and drink products for the treatment or prevention of obesity via the regulation of lipid metabolism. The organisms comprise a polynucleotide encoding a growth hormone fragment capable of stimulating the activity of hormone-sensitive lipase (the key enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key enzyme in lipogenesis). The growth hormone fragment preferably contains at least the disulphide-bonded loop of a mammalian growth hormone (but is not the full-length growth hormone) and is optionally linked to an epitope tag or heterologous fusion protein partner. The transgenic organism may be a microorganism used to produce a fermented product (e.g., yeast), or an edible plant or animal or cell thereof. Food or drink made using methods of the invention are used to modify fat/lean ratio, lipid metabolism or food use in a mammal. In particular, the food		

CC		or drink products may be used to treat or prevent obesity, particularly
CC		in humans, and may also be used to improve the fat/lean ration of
CC		livestock raised for meat production. In the exemplification of the
CC		invention, the human growth hormone (hGH) fragment analogue A09604 was
CC		expressed in yeast, optionally fused to the FLAG epitope (AAB73625).
CC		The present sequence is described as a DNA sequence from yeast in
CC		the sequence listing, but is not further referred to in the
CC		specification.
xx		
SQ		Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;
Query Match	2.3%; Score 38.8; DB 22; Length 4590;	
Best Local Similarity	15.1%; Pred. No. 1.5;	
Matches	58; Conservative 172; Mismatches 153; Indels 1; Gaps 1.	
OY	546 TTCTGTAGCGGCTGGTTTCTTTCTTAGCGGTTTTTAACGGAATAGAGTCCTTGTTGACCG	605
Dd	4050 KMCSTYYAYNHTDANDTYCTGTATTNDMGSCNHTDDDKTKKYUPTTTTDHMKDGSGDKKD	3991
OY	606 AGGGAAGAACAATGATGATCCTCTTTGCSAGGCGTTTTCTCGGTATGCTGGTTGCTACC	665
Dd	3990 CNDHDHDDMDGBVBBVBVVBBBVBBVKKCNHTDGANOHNDHNHDNGANDNDNNDKKDCX	3931
OY	666 TTTTACATTTCCCTATG-ACTGATCACTTTGATGATTAATAACAAGATCATCTCGTTGTT	724
Dd	3930 NNRBYNHNDHEDVYUNDGDGNHNDHNDHNDHNDHNBNDVWVHGSHDVYWVBYUNHNN	3871
OY	725 ATGGAATATATGCGTGAAGAATCAGGATGATCACTGAGGAAGCTTTGCTCATCTTTATC	784
Dd	3870 DHDHYUNDGDGDCAKKHMTGMRKKKKAAGMSRNKKMDSVKATKYUKKTNTCTC	3811
OY	785 TTCATCTCTGTTGATGATTAATGAGATCATGATAAAGATGTTCTTGTGGAAGAGAACAA	844
Dd	3810 TTTTASTSNNYATATMKTGTAAHNTANATAAASNSMGTDLAYCSRNYAATANATYDAKV	3751
OY	845 CATTCATCGGGAAAGTCCCTGCGTTGATTTGTTAAACAAGACATTACTTTGTTCCATCTC	904
Dd	3750 HAANKBNHTAASHNNHTDGYTSYHTFRAUDATRASVBNAKTYNNHTHAGANKBHHNTDGYTS	3691
OY	905 TGTGGTATATACCGGTTTCGATG	928
Dd	3690 YTHRAATKYCYKNSTGKNKTN SANTG	3667
RESULT 10		
ABL56202		
ID	ABL56202 standard; DNA; 50000 BP.	
XX	ABL56202;	
AC		
DT	01-JUL-2002 (first entry)	
XX		
DE	. AMEPV genome fragment#4.	
KM	AMEPV; gene therapy; viral vector; chromosome maping; gene mapping;	
KN	genetic deficiency disorder; ds.	
OS	Amsacta moorei entomopoxvirus.	
PN	WO200212526-A2.	
PD		
PF	14-FEB-2002.	
PR	10-AUG-2001; 2001WO-US25287.	
XX		
PR	10-AUG-2000; 2000US-224479P.	
XX	14-SEP-2000; 2000US-0662254.	
PA	(UYFL) UNIV FLORIDA.	
PI	Moyer RW, Li Y, Bawden AL,	
DR	WPI; 2002-227161/28.	

XX Novel recombinant entomopox virus vector useful for delivering
PT polynucleotide encoding protein to vertebrate cell, comprises
PT polynucleotide encoding protein operably linked with heterologous
PT promoter sequence -
XX
PS Disclosure; Page 201-226; 326bp; English.
XX
CC The invention relates to a recombinant entomopox virus (EPV) vector,
CC comprising a polynucleotide encoding a protein operably linked with a
CC heterologous promoter sequence. The invention also concerns methods for
CC providing gene therapy for genetic deficiency disorders. Vectors of the
CC invention are useful for delivering a polynucleotide encoding a protein
CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC The vector is introduced into the vertebrate cell by infection in a viral
CC particle, or by transfection, transduction, or injection either in vitro
CC or in vivo. The vector is useful for the delivery and expression of
CC biologically useful proteins in gene therapy protocols, and for
CC delivering large DNA segments for engineering of vertebrate cells.
CC Polynucleotides of the invention have applications in techniques such as
CC their use as insertion sites for foreign genes of interest, hybridisation
CC probes, for chromosome and gene mapping, in PCR technologies, and in the
CC production of sense or antisense nucleic acids. Vectors of the invention
CC provide for stable integration and expression of heterologous DNA in host
CC cells, and are adapted for accepting large heterologous polynucleotide
CC inserts which can be delivered in an infected or transformed cell and
CC expressed in a stable fraction. The current sequence represents a
CC fragment of the genome of the genus B entomopoxvirus from *Amsacta moorei*
CC (AmpVY).
XX
SQ Sequence 50000 BP; 20248 A; 4709 C; 4703 G; 20340 T; 0 other;
XX
Query Match 2.3%; Score 38.2; DB 24; Length 50000;
Best Local Similarity 52.1%; Pred. No. 7.1;
Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
XX
QY 641 TTCTCGTAGTGTGCTGCTACCTTTAGATTCCCTATGACTGATCAAGTTGATGAT 700
DB 48598 TTCTCTATTATTATTATTGTTTTTATTATTATGATATTAAATTTTCATTTT 48657
XX
QY 701 TAAATCAGATCATCTGCTGTTATGATATATGATGGAAGATCGGATGATGATCG 760
DB 48658 AACATTAAATTAATTTTCGTCATTATATTTTGAATATATTAAGATTAATCATCAAT 48717
XX
QY 761 AGGGAACCTTGCTCATCTTTATCTTCATCTTGTCATGATTA 803
DB 48718 TAAATATTAATTAATTAATTAACATTCATCAATTTTATATTA 48760
XX
RESULT 11
AAS62336/c
ID AAS62336 standard; cDNA; 1962 BP.
XX
AC AAS62336;
XX
DT 14-FEB-2002 (first entry)
XX
DE cDNA sequence #123 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antineumatic; ss.
XX
OS Homo sapiens.
XX
PN WO200177291-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10485.
XX
PR 06-APR-2000; 2000US-195604P.

XX
PA (GENM) GENNETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
DR WPI; 2002-010900/01.
XX
PT New polynucleotides encoding secreted proteins useful for treating e.g.
PT asthma, HIV and Crohn's disease -
XX
PS Claim 1; Page 144; 391pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides
CC a method for producing proteins from these polynucleotide sequences.
CC The proteins are useful for identifying compounds that modulate their
CC activity and production, and the cell is also useful for identifying
CC compounds that modulate expression of the polynucleotide sequences
CC encoding the secreted proteins. The sequences of the invention are
CC useful for treating diseases such as hyperproliferative disorders
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined
CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
CC The polynucleotide sequences of the invention are also useful in gene
CC therapy. AAS62214-AAS6238 represent the cDNA sequences of the
CC invention that encode for novel human secreted proteins.
XX
SQ Sequence 1962 BP; 620 A; 340 C; 320 G; 682 T; 0 other;
XX
Query Match 2.3%; Score 37.8; DB 24; Length 1962;
Best Local Similarity 58.4%; Pred. No. 2;
Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
XX
QY 708 AGAATCATCTGCTGTTTATGATATATGATGAAGATCAGTGTGATGAGGAGAC 767
DB 668 AGAATATCAGGTTTATATCTTGATCTTTCATGAACAGCTTGTGACACTGAAGAT 609
XX
QY 768 TTGTCATCATCTTATVCTTCATCTGTCATGATTAAGAGATCATGATAGA 820
DB 608 CTACTAAATCTTATGATTAAATGTTCTGCTGTGAACCTGAATAATAGA 556
XX
RESULT 12
AAS63335
ID AAS63335 standard; cDNA; 2839 BP.
XX
AC AAS63335;
XX
DT 19-APR-2001 (first entry)
XX
DE Human GTP-binding associated protein #35 coding sequence.
XX
KW Human; guanosine triphosphate binding associated protein; GTP; GAP;
KW inflammation; AIDS; Addison's disease; anemia; arteriosclerosis; asthma;
KW autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;
KW osteoporosis; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200105970-A2.
XX
PD 25-JAN-2001.
XX
PF 19-JUL-2000; 2000WO-US19698.
XX
PR 19-JUL-1999; 99US-0144595.
PR 23-AUG-1999; 99US-0150460.
PR 15-OCT-1999; 99US-0159849.
XX
PA (INCY-) INCYTE GENOMICS INC.

```
XX Yue H, Tang YT, Bandman O, Hillman JL, Lal P, Au-Young J;
PI Reddy R, Yang J, Baughn MR, Lu DAM, Azimzal Y, Patterson C;
XX WPI: 2001-091972/10.
DR P-PSDB; AAB68535.
XX
XX New guanosine triphosphate-binding associated proteins (GBAP) and their
PT encoding nucleic acids, useful for treating and/or diagnosing diseases
PT associated with GBAP expression, such as cancer, diabetes and asthma -
XX
XX Example 5; Pages 210-211; 233pp; English.
XX
XX The present invention relates to novel human guanosine triphosphate
CC (GTP)-binding associated proteins (GBAPs; AAB68501-AAB68566) and their
CC coding sequences (AAE58301-AAE58365). The proteins and coding sequences
CC of the present invention are useful for treating a variety of disorders
CC including inflammation, AIDS, Addison's disease, anaemia,
CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,
CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
CC psoriasis.
XX
SQ Sequence 2839 BP; 894 A; 493 C; 654 G; 798 T; 0 other;

Query Match      2.3%; Score 37.8; DB 22; Length 2839;
Best Local Similarity 50.8%; Pred.No.2.4;
Matches 90; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

OY 675 CCCATGACTGATCGATTGATGATTAATCAAGATCATCTGTTGATGATATAT 734
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2512 CACAAATGATTTGGCAGTTTCTTTGATTCATTAATTAACATGTTTAAGAAAA 2571
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 735 GGTGAAGATCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 794
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2572 GTAACCTAATTTGGGTTTTCAGGCGAGTTGATTAATGACCTTAAGCTTTT 2631
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 795 TCATGATTATGAGATCATGATGATGATGATGATGATGATGATGATGATGAT 851
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2632 TTTTTTTTCAGATTAAATGCTAAGAAAAGATTTGGGAGAGTTAATTAAGTAT 2688
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
AAS42103/C
ID AAS42103 standard; DNA; 9516 BP.
XX
XX AAS42103;
AC
XX
XX 17-DEC-2001 (first entry)
DT
XX
XX Genomic sequence #419 encoding novel human enzyme polypeptide.
DE
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KM ligase; hyperproliferative disorder; immunodeficiency disorder;
KM autoimmune disorder; neurological disorder; metabolic disorder;
KM inflammatory disorder; cardiovascular disorder; reproductive disorder;
KM blood-related disorder; infectious disorder; gene therapy; cytostatic;
KM anti arthritic; nephrotropic; anticoagulant; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200155301-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01239.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184654.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
```

```
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214866.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0237935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
```


XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX
SQ Sequence 6392 BP; 1818 A; 170 C; 1388 G; 3016 T; 0 other;

Query Match 2.3%; Score 37.6; DB 24; Length 6392;
Best Local Similarity 52.6%; Pred. No. 4;

Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 669 AGATTTCCTATGACATCAGTTGATGATTAAATCAAGATCATCTCGTTGTTATGG 728

DB 3646 ATATTTTCTTATGAGATGAGAAATTTTGTGTAGTTGATTGATTAATGGTTGTTAA 3705

QY 729 ATATATGTCGAGAAATCAGTGTATGATCTGAGGAGACTTGTCTCATCTTATCTTCA 788

DB 3706 TTATATAGTCGATGATTTGTTTTCATTTCCGGGGTTTTTTTATTTATTTAGTTT 3765

QY 789 TCTTGTTCATGATTAATGAGATCAGATTAAGATGTT 824

DB 3766 TTTTATTTTAAATGAAGAAAGTAAGATTTGGTGT 3801

RESULT 15

ABL34507
ID ABL34507 standard; DNA; 6392 BP.

XX ABL34507;

DT 26-MAR-2002 (first entry)

XX Human metastasis associated gene SEQ ID NO: 60.

XX Metastasis associated gene; cytostatic; gene therapy; cancer;

XX cytosine methylation; gene; ds.

XX Homo sapiens.

XX WO200177376-A2.

XX 18-OCT-2001.

XX 06-APR-2001; 2001WO-EP03970.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-010922/01.

XX New nucleic acid derived from chemically treated metastasis genes,
XX useful for diagnosis of cancers by analysis of cytosine methylation,
XX also for treatment -

XX Claim 1; SEQ ID NO 60; 23pp + Sequence Listing; English.

XX The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention.

XX Sequence 6392 BP; 1818 A; 170 C; 1388 G; 3016 T; 0 other;

Query Match 2.3%; Score 37.6; DB 24; Length 6392;
Best Local Similarity 52.6%; Pred. No. 4;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 669 AGATTTCCTATGACATCAGTTGATGATTAAATCAAGAAATCTCGTTGTTATGG 728

DB 3646 ATATTTTCTTATGAGATGAGAAATTTTGTGTAGTTGATTGATTAATGGTTGTTAA 3705

QY 729 ATATATGTCGAGAAATCAGTGTATGATCTGAGGAGACTTGTCTCATCTTATCTTCA 788

DB 3706 TTATATAGTCGATGATTTGTTTTCATTTCCGGGGTTTTTTTATTTATTTAGTTT 3765

QY 789 TCTTGTTCATGATTAATGAGATCAGATTAAGATGTT 824

DB 3766 TTTTATTTTAAATGAAGAAAGTAAGATTTGGTGT 3801

Search completed: March 5, 2003, 10:46:31

Job time : 560 secs

GenCore version 5.1.4_P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 5, 2003, 10:35:30 ; Search time 88 Seconds
(without alignments)
5792.009 Million cell updates/sec

Title: US-10-037-311A-2

Perfect score: 1662

Sequence: 1 atggatcgaattcgtacag.....ggggacttaagctagatga 1662

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.6	2.6	7218	1	US-08-232-463-14 Sequence 14, Appl
2	37	2.2	15225	2	US-08-892-403A-2 Sequence 2, Appl1
3	35.4	2.1	2935	4	US-09-480-921B-27 Sequence 27, Appl
4	35.4	2.1	7218	1	US-08-232-463-14 Sequence 14, Appl
5	35.4	2.1	8050	4	US-09-491-362-11 Sequence 11, Appl
6	35.4	2.1	8050	4	US-09-874-562-11 Sequence 11, Appl
7	34	2.0	87350	3	US-08-781-891-79 Sequence 79, Appl
8	34	2.0	87543	4	US-09-791-211-3 Sequence 3, Appl1
9	33.4	2.0	14872	1	US-08-961-527-72 Sequence 72, Appl
10	32.6	2.0	7400	3	US-08-473-185-1 Sequence 1, Appl
11	32.6	2.0	7400	3	US-09-171-387-3 Sequence 1, Appl
12	32.6	2.0	7400	3	US-09-853-768-10 Sequence 3, Appl1
13	32.4	1.9	5852	4	US-09-853-768-3 Sequence 10, Appl
14	32.4	1.9	7037	4	US-09-853-768-3 Sequence 3, Appl1
15	32.2	1.9	4190	2	US-08-488-706-3 Sequence 3, Appl1
16	32.2	1.9	4450	3	US-08-617-860B-2 Sequence 2, Appl1
17	32.2	1.9	2364	2	US-08-838-219B-5 Sequence 5, Appl1
18	32.2	1.9	2364	3	US-09-233-336A-5 Sequence 5, Appl1
19	32.2	1.9	2364	3	US-09-233-336A-5 Sequence 5, Appl1
20	32.2	1.9	2364	4	US-09-402-036-5 Sequence 5, Appl1
21	32.2	1.9	2364	4	US-09-904-226-5 Sequence 5, Appl1
22	32.2	1.9	2612	1	US-08-471-023-31 Sequence 31, Appl
23	32.2	1.9	2612	2	US-08-471-044-31 Sequence 31, Appl
24	32.2	1.9	2612	2	US-08-463-483A-31 Sequence 31, Appl
25	32.2	1.9	2612	2	US-08-471-046A-31 Sequence 31, Appl
26	32.2	1.9	2612	2	US-08-470-566B-31 Sequence 31, Appl
27	32.2	1.9	2612	2	US-08-838-219B-3 Sequence 3, Appl1

C 28	32	1.9	2612	2	US-08-469-334-31	Sequence 31, Appl
C 29	32	1.9	2612	3	US-09-300-529-31	Sequence 31, Appl
C 30	32	1.9	2612	3	US-09-233-336A-3	Sequence 3, Appl1
C 31	32	1.9	2612	3	US-09-233-752A-3	Sequence 3, Appl1
C 32	32	1.9	2612	4	US-09-402-036-3	Sequence 3, Appl1
C 33	32	1.9	2612	4	US-09-904-226-3	Sequence 3, Appl1
C 34	32	1.9	3472	4	US-09-678-300-14	Sequence 14, Appl
C 35	31.6	1.9	1074	4	US-09-134-001C-1643	Sequence 2054, Ap
C 36	31.6	1.9	2022	4	US-09-134-001C-1643	Sequence 1643, Ap
C 37	31.6	1.9	2746	2	US-08-576-165-3	Sequence 23, Appl
C 38	31.2	1.9	218	1	US-08-700-575-23	Sequence 718, App
C 39	31.2	1.9	544	4	US-09-328-111-718	Sequence 5, Appl1
C 40	31.2	1.9	792	3	US-08-565-386-5	Sequence 1, Appl1
C 41	31.2	1.9	1266	3	US-09-190-821-1	Sequence 1, Appl1
C 42	31.2	1.9	1266	4	US-09-535-493-1	Sequence 4, Appl1
C 43	31.2	1.9	1437	4	US-09-137-222A-4	Sequence 1, Appl1
C 44	31.2	1.9	4529	1	US-08-565-386-1	Sequence 1, Appl1
C 45	31.2	1.9	4739	3	US-08-685-871-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14
Query Match 2.6%; Score 43.6; DB 1; Length 7218;

1	APPLICANT:	Donna T. Ward
2	APPLICANT:	Andrew T. Wallt
3	TITLE OF INVENTION:	ANTISENSE MODULATION OF WRN EXPRESSION
4	FILE REFERENCE:	RTS-0205
5	CURRENT APPLICATION NUMBER:	US/09/791,211
6	CURRENT FILING DATE:	2001-02-23
7	NUMBER OF SEQ ID NOS:	90
8	SEQ ID NO 3	
9	LENGTH:	87543
10	TYPE:	DNA
11	ORGANISM:	Homo sapiens
12	FEATURE:	
13	NAME/KEY:	unsure
14	LOCATION:	7421
15	OTHER INFORMATION:	unknown
16	NAME/KEY:	unsure
17	LOCATION:	7427
18	OTHER INFORMATION:	unknown
19	NAME/KEY:	unsure
20	LOCATION:	11609
21	OTHER INFORMATION:	unknown
22	NAME/KEY:	unsure
23	LOCATION:	12605
24	OTHER INFORMATION:	unknown
25	NAME/KEY:	unsure
26	LOCATION:	12742
27	OTHER INFORMATION:	unknown
28	NAME/KEY:	unsure
29	LOCATION:	29370
30	OTHER INFORMATION:	unknown
31	NAME/KEY:	unsure
32	LOCATION:	29422
33	OTHER INFORMATION:	unknown
34	NAME/KEY:	unsure
35	LOCATION:	29979
36	OTHER INFORMATION:	unknown
37	NAME/KEY:	unsure
38	LOCATION:	29980
39	OTHER INFORMATION:	unknown
40	NAME/KEY:	unsure
41	LOCATION:	29981
42	OTHER INFORMATION:	unknown
43	NAME/KEY:	unsure
44	LOCATION:	30136
45	OTHER INFORMATION:	unknown
46	NAME/KEY:	unsure
47	LOCATION:	30140
48	OTHER INFORMATION:	unknown
49	NAME/KEY:	unsure
50	LOCATION:	31205
51	OTHER INFORMATION:	unknown
52	NAME/KEY:	unsure
53	LOCATION:	31206
54	OTHER INFORMATION:	unknown
55	NAME/KEY:	unsure
56	LOCATION:	31592
57	OTHER INFORMATION:	unknown
58	NAME/KEY:	unsure
59	LOCATION:	33095
60	OTHER INFORMATION:	unknown
61	NAME/KEY:	unsure
62	LOCATION:	33160
63	OTHER INFORMATION:	unknown
64	NAME/KEY:	unsure
65	LOCATION:	34066
66	OTHER INFORMATION:	unknown
67	NAME/KEY:	unsure
68	LOCATION:	34072
69	OTHER INFORMATION:	unknown
70	NAME/KEY:	unsure
71	LOCATION:	36816
72	OTHER INFORMATION:	unknown
73	NAME/KEY:	unsure

LOCATION: 39020
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42164
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42459
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46808
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46823
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46826
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 47291
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52786
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52787
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 53384
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 54684
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59215
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59235
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59242
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 63290
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 66614
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68660
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68697
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68718
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68733
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68739
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 69785
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79134
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79198
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 86336

OTHER INFORMATION: unknown
OTHER INFORMATION:
US-09-791-211-3

Query Match 2.0%; Score 34; DB 4; Length 87543;
Best Local Similarity 53.8%; Pred. No. 13;
Matches 70; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 283 GGGCTACTGCTTCGCTTGATGAGATTCCTGCTTAGTAGGACCAATCAGTCTCAT 342
DB 5379 GTGATTTTACCTCATATTATTTACGAGATGACCTTAACATATACGTTTCGTTCTTTTAT 5438
QY 343 TACCGTAACCTTCACCTTACAGCCATCTTATCTCATCTTAAGCTTAGAACTAC 402
DB 5439 TAATTAATAATTTGCTTCGGTAGATCATCTGCTCATTTTGGACAAACAGTAC 5498
QY 403 GAAAGCTTC 412
DB 5499 AAAAGCCTC 5508

RESULT 9

US-08-961-527-72
Sequence 72, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 14872 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-72

Query Match 2.0%; Score 33.4; DB 4; Length 14872;
Best Local Similarity 58.6%; Pred. No. 7.8;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1232 CATCTTTGAACGGGGTTACGGGAGAACTTAAGAGTATGTTGGAAATATCCACAT 1291
DB 14327 CATCTTGCAGAAAGAGAGCTCGGAGATTTTAAGTAATAATTTAGTAATGAAGAAGAA 14386
QY 1292 CAACGTGAGAAATCATCGGTTCATCAGCCGAGCCAG 1330
111 1111111 11 1111 1 111

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 10:42:45 ; Search time 160 Seconds
(without alignments)
6938.018 Million cell updates/sec

Title: US-10-037-311a-2
Perfect score: 1662
Sequence: 1 atgacatcagatcgtacag.....ggggacttaagctatgta 1662

Scoring table: JIDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 478924 seqs, 33395956 residues

Total number of hits satisfying chosen parameters: 957848

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA :
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCF_NEM_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCFUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	670.4	40.3	1614	9	US-09-938-842A-1989
2	606	36.5	1512	9	US-09-938-842A-1990
3	113.2	6.8	382	10	US-09-878-574-754
4	50.6	3.0	171	10	US-09-294-0938-5456
5	47.6	2.9	228	10	US-09-923-876-5543
6	39.2	2.4	545	10	US-09-878-574-4288
7	37.8	2.3	1962	10	US-09-822-830A-123
8	36.6	2.2	1022	10	US-09-770-445-210
9	36	2.2	328	10	US-09-924-035A-534
10	36	2.2	639	10	US-09-770-149-623
11	35.4	2.1	944	10	US-09-974-300-2061
12	35.4	2.1	2935	9	US-10-047-412A-27
13	35.4	2.1	3400	10	US-09-987-025-1
14	35	2.1	229	10	US-09-969-373-1042
15	35	2.1	660	10	US-09-733-507-3
16	35	2.1	904	10	US-09-733-507-1
17	34.8	2.1	30568	10	US-09-764-877-3851
18	34.8	2.1	172637	10	US-09-805-458A-3
19	34.6	2.1	16281	10	US-09-764-847-1367

20	34.6	2.1	16285	10	US-09-764-847-1368	Sequence 1368, Ap
21	34.6	2.1	16285	10	US-09-764-847-1369	Sequence 1369, Ap
22	34.6	2.1	684973	10	US-09-263-959-1	Sequence 1, Appl1
23	34.4	2.1	372	10	US-09-960-352-6252	Sequence 6252, Ap
24	33.8	2.0	813	9	US-09-938-842A-1178	Sequence 1178, Ap
25	33.2	2.0	1141	12	US-10-007-693-41	Sequence 41, Appl1
26	33.2	2.0	2162	9	US-09-895-913A-3	Sequence 3, Appl1
27	33	2.0	3596	10	US-09-887-576-451	Sequence 451, Ap
28	33	2.0	5190	10	US-09-070-927A-395	Sequence 395, Ap
29	33	2.0	5222	10	US-09-416-384A-1	Sequence 1, Appl1
30	32.8	2.0	410	10	US-09-864-761-1209	Sequence 1209, Ap
31	32.8	2.0	978	9	US-09-938-842A-2911	Sequence 2911, Ap
32	32.8	2.0	1035	9	US-09-938-842A-1734	Sequence 1734, Ap
33	32.8	2.0	2000	9	US-09-938-842A-822	Sequence 4822, Ap
34	32.8	2.0	2004	10	US-09-887-576-320	Sequence 320, Ap
35	32.6	2.0	603	10	US-09-864-761-18328	Sequence 18328, A
36	32.6	2.0	914	10	US-09-864-761-32148	Sequence 32148, A
37	32.6	2.0	964	10	US-09-864-761-1570	Sequence 1570, Ap
38	32.6	2.0	1964	10	US-09-864-761-15641	Sequence 15641, A
39	32.6	2.0	6107	9	US-09-984-271-47	Sequence 47, Appl1
40	32.4	1.9	2004	10	US-09-887-576-178	Sequence 178, Ap
41	32.4	1.9	5775	9	US-09-866-557A-1	Sequence 1, Appl1
42	32.2	1.9	2000	9	US-09-938-842A-5145	Sequence 5145, Ap
43	32.2	1.9	3336	10	US-09-815-242-8857	Sequence 8857, Ap
44	32.2	1.9	99014	10	US-09-880-107-3428	Sequence 3428, Ap
45	32	1.9	429	10	US-09-860-432-58	Sequence 58, Appl1

ALIGNMENTS

RESULT 1
US-09-938-842A-1989
; Sequence 1989, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938, 842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227, 866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264, 647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300, 111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1989
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1989
Query Match 40.3%; Score 670.4; DB 9; Length 1614;
Best Local Similarity 69.4%; Pred. No. 4.1e-185;
Matches 962; Conservative 0; Mismatches 406; Indels 18; Gaps 3;
QY 268 GATACGCTTCGGAGCGCTACTGCTTGGTTTGATGAAGATTCTTCCTAGTAGG 327
DB 190 GATAGCATGATCGAGGCGCTTAAACCGGGAATTTCGATGAAGGCTTCCTGATGATAGG 249
QY 328 TGC---CAATGATTCATTAACCGTAAACCTTACACCTTACAGCATCTTATCTATCAGC 384
DB 250 TATACAGCAATCTTTGTTGATCCGGAAGGCTTACACCGCTTTCGGAATATCTTGTGC 309
QY 385 TCTAAGCTTAGAAGTACGAAGGCTTCAAGCGATGTGTCGGGCTACTGATATCTTAC 444
DB 310 TCTAAGCTTAGAAGTACGAAGGCTTCAAGCGTTCACAGCGTTGGCGGTCCAGGACGATGCTTAC 369


```
Db 421 GATAGATATCTTCTTACAAACCGCAAGATATGCGATCTCTTATGCGAGCATTT 480
OY 643 CTCGGATGTCGGGGTGTACCTTATAGATTCCTTATGCGATGATGATGATTA 702
Db 481 CCAAGTACTTCAATGGTGGTCTCTCGACTTCCATTTATGAAATATGCTGTGGATAC 540
OY 703 AATCAAGATCACTCGTGTATGATATATGATGATGATGATGATGATGATGATGATG 762
Db 541 CAAAGGATGATCTGCTGTGTACGSAACAATGTTGAAATATCATTCATCAACTGACT 600
OY 763 GCAACTTGTCTCATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTAT 822
Db 601 TCATTCCTCCCACTATATATGATATGATATGATATGATATGATATGATATGATG 660
OY 823 TTTCTCTGAGGAGACCAATCATCGGGAAGTCCCTGTTGATTTGTTAAAGA 882
Db 661 TTTCTCTGAGGAGATCAAGATTTGATGACAAAGTCCCTGTTGATTTGATGAGCC 720
OY 883 GACAAATTAATCTTCTCATCTCTGTTTAAATACCGGTTTGCATGATGATGATGATG 942
Db 721 AATGTTTACTTGTTCATCTTATGATTTATGATTTATGATTTATGATTTATGATG 780
OY 943 CTATTCCTCAAGAAACGACTCTTTCATCTTATGATGATGATGATGATGATGATG 1002
Db 781 CTGTTCCCAAGAAACGCTGTTTACCACTTGGTGGTATCTTTCACCTTAA 840
OY 1003 AACCAATTAATGAGGCTTACTGATGATGATGATGATGATGATGATGATGATGATG 1062
Db 841 AATCAAGTGTGAGATATGCTCACTAATGATGATGATGATGATGATGATGATGATG 900
OY 1063 AAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1122
Db 901 AGACTGGGATTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
OY 1123 GATCAGATTTCTTCTTATCTTATGATGATGATGATGATGATGATGATGATGATG 1182
Db 961 GACCAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
OY 1183 GAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1239
Db 1021 GATCAAAAGTCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
OY 1240 AACGGGTTTACGCGAGAACTTAAAGATATGATGATGATGATGATGATGATGATG 1299
Db 1081 TCTCCTGATCTCTAAGAACTAGAACTATGATGATGATGATGATGATGATGATG 1140
OY 1300 GAAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1359
Db 1141 GAGATTTATTAAGTATATGATGATGATGATGATGATGATGATGATGATGATG 1200
OY 1360 CATTAATGGAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1419
Db 1201 CACGACAAAAGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 1260
OY 1420 ACTGCTTGTCTATATGATGATGATGATGATGATGATGATGATGATGATGATG 1479
Db 1261 ACCTCAAGTCAACATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1320
OY 1480 CTCATATGACCGGAAACCGTACAACTCCGATCTTCTGCTGCTGCTGCTGCTGCT 1539
Db 1321 CTTTATCTGCTCAAAATGATTAACAAAGCTCTGATCCGCTGCTGCTGCTGCTG 1380
OY 1540 GAGCCTTGTCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1596
Db 1381 GAGCCTTGTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
OY 1597 ACAGGAACACTAGTCTCATGATGATGATGATGATGATGATGATGATGATGATG 1644
Db 1441 TCGGGGAAGTGTCTTCTTCTGATGATGATGATGATGATGATGATGATGATG 1488
```

RESULT 3

```
US-09-878-574-754
; Sequence 754, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 754
; LENGTH: 382
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-046-Q1-B1-H12
US-09-878-574-754

Query Match
Best Local Similarity 57.9%; Pred. No. 4.9e-23;
Matches 221; Conservative 0; Mismatches 158; Indels 3; Gaps 1;

OY 944 TATTCACAGAAAGGAGCTGCTTCTCATCTAGTATGATGATGATGATGATGATGATG 1003
Db 1 TGTTCACAGAGAGACATTTATTTCCACCTTGGTGGTCTGCTGCTGCTGCTGCTGCTGCT 60
OY 1004 ACCAAGTATGAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1063
Db 61 AGCATGATGAGGAGAACTATATGATGATGATGATGATGATGATGATGATGATGATG 120
OY 1064 AGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1123
Db 121 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
OY 1124 ATCAATTTCTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1180
Db 181 ATCTAGTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
OY 1181 TGGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1240
Db 241 CAGTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
OY 1241 ACAGGCTTACCGGAGAACTTAAAGATGATGATGATGATGATGATGATGATGATG 1300
Db 301 ATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
OY 1301 AATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1322
Db 361 AAGTCATTAAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 382

RESULT 4
US-09-294-093B-5456
; Sequence 5456, Application US/09294093B
; Patent No. US2001005135A1
; GENERAL INFORMATION:
; APPLICANT: Lalpudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 5456
; LENGTH: 171
```



```
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 123
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-123
```

```
Query Match          2.3%; Score 37.8; DB 10; Length 1962;
Best Local Similarity 58.4%; Pred. No. 1.1;
Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
```

```
QY 708 AGAATCATCTGTTTATGATATATGATGGAAGATCAGTATGATGAGGAAC 767
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 668 AGAATATCCAGTTAATCTTGATCTTTCATGAAACAGCTTTGACACCTGAAGAA 609
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 768 TTGTCTCATCTTATCTTCATCTTCTTCATGATTTATGAGATCATGATAGA 820
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 608 CTACTAATCTTCTTATGATTTATGTTCTGTCTGTAAGCTGAAATATAGA 556
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

RESULT 8

```
US-09-770-445-210/c
; Sequence 210, Application US/09770445
; Patent No. US2002023281A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Ted
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210
; LENGTH: 1022
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-210
```

```
Query Match          2.2%; Score 36.6; DB 10; Length 1022;
Best Local Similarity 68.0%; Pred. No. 1.8;
Matches 51; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
```

```
QY 421 TGTGTCGCGGTACTGATCTTACAGAAAGCTTTAAACCACTTATCAAGACATATT 480
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 807 TGTAGTGGAGCAATGAAATATAGAGAAAGATTAATACATCTGGAGAGGAAGATATA 748
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 481 GATGTGATGTGTA 495
      |||||  |||||  |||||
DB 747 GATGTGACACTGAA 733
      |||||  |||||  |||||
```

```
RESULT 9
US-09-924-035A-534/c
; Sequence 534, Application US/09924035A
; Patent No. US20020142319A1
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2011US
; CURRENT APPLICATION NUMBER: US/09/924,035A
; CURRENT FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,784
; PRIOR FILING DATE: 1999-08-13
; NUMBER OF SEQ ID NOS: 900
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 534
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-924-035A-534
```

```
Query Match          2.2%; Score 36; DB 10; Length 328;
Best Local Similarity 58.3%; Pred. No. 1.4;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
QY 1310 GTGTCTCAGCCGACCAAGAGTTATCAGCAGACCGAAAAAGATCATATGCGCA 1369
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 312 GATTTGTCGTCGTGAGCAGCAAGATCATGATGATCTTACGATCTTATGATTTCC 253
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1370 AAGCTCTGCGGAATGTATCTTTTGTGATTTGACAGATATATCTTGGA 1417
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 252 ATGCTTAAGCGGCGCTGTCATGTGGTTGACTTATCATGTTTGA 205
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

RESULT 10

```
US-09-770-149-623/c
; Sequence 623, Application US/09770149
; Patent No. US2002005963A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Ted
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 623
; LENGTH: 639
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
```

```
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION:(1)...(639)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-149-623

Query Match          2.2%; Score 36; DB 10; Length 639;
Best Local Similarity 58.3%; Pred No. 2.1;
Matches   63; Conservative    0; Mismatches   45; Indels      0; Gaps      0;
```

```
OY       1310 GTGTCATCAGCCCGACCAAGAAGTTTATCACACGCCAAMAAAAGATGCATAATGGCA 1369
           ||| || | ||||| ||||| || || || || | || | ||| ||| |||
DB        300 GAGTGCTGTGGTGAGACCACAAGCATGACTGTCTCATCTTAACGAGATCTGTATCCC 241
           ||| ||| ||||| || || || || || || || || || || || || ||

OY       1370 AAGCTTTCGCCGAANGTATTCTTTGACGTTTGACAGATATCTTTGGA 1417
           ||| ||| ||||| || || || || || || || || || || || || ||
DB        240 ATGCTTAAGCGGGCTGTGTCATTGTGGGTTGACTTATCATCATGTTTTTA 193
           ||| ||| ||||| || || || || || || || || || || || || ||
```

```
RESULT 11
US-09-974-300-2061
; Sequence 2061, Application US/099743300
; Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085-500-US
CURRENT FILING DATE: 2001-09/974,300
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: PASTSEQ for Windows Version 4.0
SEQ ID NO 2061
LENGTH: 944
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-2061
```

```
Query Match          2.1%; Score 35.4; DB 10; Length 944;
Best Local Similarity 53.2%; Pred. No. 3.8;
Matches   75; Conservative    0; Mismatches   66; Indels      0; Gaps      0;
```

```
OY       1243 GCGGGTTACGCGGAGAACTTAAGAGTAGTATTGGGAATATCCGACATCACTGAGGAA 1302
           ||| ||| ||||| || || || || || || || || || || || || ||
DB        421 GAGCTTGCACGAAGAAGACTACATGACAGCGTTTAAAGAAAAATFACGGAAAAACATTGGAAAA 480
           ||| ||| ||||| || || || || || || || || || || || || ||

OY       1303 ATCATCGGTGTTTCATCAGCCGACCAAGAGAGTTATCAGCAGACCGAAAAAAGATGCAT 1362
           ||| ||| ||||| || || || || || || || || || || || || ||
DB        481 ATTCTTGGCAAAGAAAGATGCTGTCAAAAGAACCCCTTGGAATAATCGAATAAGTAGAG 540
           ||| ||| ||||| || || || || || || || || || || || || ||

OY       1363 AATGGCAAAAGCTCTTGGCGAA 1383
           ||| ||| || || || ||
DB        541 AAAGCCAACAAGCTGGCGAAA 561
           ||| ||| || || || ||
```

```
RESULT 12
US-10-047-412A-27
; Sequence 27, Application US/10047412A
; Publication No. US20020197696A1
GENERAL INFORMATION:
APPLICANT: Levin, Joshua Z.
APPLICANT: Budziszewski, Gregory J.
APPLICANT: Potler, Sharon L.
APPLICANT: Wegrich, Lynette M.
TITLE OF INVENTION: Herbicide Target Genes and Methods
FILE REFERENCE: PB/5-30780DIV
CURRENT APPLICATION NUMBER: US/10/047,412A
CURRENT FILING DATE: 2002-04-11
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.1
```

```
; SEQ_ID NO 77
; LENGTH: 2935
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-047-412A-27

Query Match          2.1%; Score 35.4; DB 9; Length 2935;
Best Local Similarity 50.3%; Pred. No. 7;
Matches      87; Conservative      0; Mismatches    86; Indels      0; Gaps      0
```

```

RESULT 13
US-09-987-025-1/c
; Sequence 1, Application US/09987025
; Patent No. US20020108148A1
GENERAL INFORMATION:
APPLICANT: Boronati, Albert
APPLICANT: Campos, Narcisco
APPLICANT: Kishore, Ganesh M.
TITLE OF INVENTION: Nucleic Acid Sequences Involved In
FILE REFERENCE: 17142/02/US
CURRENT FILING DATE: 2001-11-13
PRIORITY FILING DATE: 09/549,787
PRIOR APPLICATION NUMBER: 60/129,899
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: 60/146,461
PRIOR FILING DATE: 1999-07-30
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3400
TYPE: DNA
ORGANISM: Arabidopsis sp
US-09-987-025-1

Query Match      2.1%; Score 35.4; DB 10; Length 3400;
Best Local Similarity 50.3%; Pred. No. 7.6;
Matches 87; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 457 AACACACTGGATCAGACGATATTGTGATGCGTGAATGCAGAAATATGTTGTGGATT 516
    ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 2335 AACCAATTTGAGCAAGCACACATGATCTCTTGACACAATAAAATACATGTTATACAGT 2276

QY 517 TCTTTTAGCGGCTTAAGGACAGCATCTTCCTAGCCCTGGTTTTCTTTACCGCCTT 576
    |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 2275 TATTTTAAAAACCGGAAAATATATATATATATAGTTAGTAAGTTTCACAGACCGTGCT 2216

QY 577 TTACAGGATTAAGTCTTCTGTTGACCAGAGGAAAGACATGATGATCTCTT 629
    ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 2215 TTCATCATATGGAATGATGATCTTTGGCGGATGAATGACAAATCTCTATATGCT 2163

RESULT 14
US-09-969-373-1042/c
; Sequence 1042, Application US/09969373
; Patent No. US2002013852A1
GENERAL INFORMATION:
APPLICANT: Eiferitz, Roger J.
```

```

; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 1042
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-1042

Query Match
Best Local Similarity 53.2%; Score 35; DB 10; Length 229;
Matches 74; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

OY 689 AGTTGATGATTAATCAAGATCATCTGTTGTTATGATATATGAGATCAGG 748
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 180 AATTAAATTAATTAAGTAATAACATTAATTAATGCTCATTAATTAATAA 121

OY 749 TGATTGATCTGAGGACTTGTCTCATCTTATCTCATCTGTCATGATTAAG 808
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 120 TTAATCATTAATTTTAACTGTTTATCTCATTTTGAATTTGTTAATTTCTTAT 61

OY 809 ATCATGATAAGATGTTCTT 827
    | | | | | | | | | |
Db 60 ACTAATAATAATATTTTATT 42

RESULT 15
US-09-733-507-3
; Sequence 3, Application US/09733507
; Patent No. US20010025379A1
; GENERAL INFORMATION:
; APPLICANT: Agriculture and Agri-Food Canada; The University of Saskatchewan
; TITLE OF INVENTION: Cyclicin Dependant Kinase Inhibitors as Plant Growth
; FILE REFERENCE: 81601-3
; CURRENT APPLICATION NUMBER: US/09/733,507
; CURRENT FILING DATE: 2000-12-02
; PRIOR APPLICATION NUMBER: CA 2,256,121
; PRIOR FILING DATE: 1998-12-31
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 660
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-733-507-3

Query Match
Best Local Similarity 66.7%; Score 35; DB 10; Length 660;
Matches 50; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

OY 421 TGTGTCGCGGACTGATCTTACAGAAAGCTCTAAACAACCTGATCAGAACATATT 480
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 TGTAGTGAAGCAATGATATAAGAAAGAAATTAATACATCTGAGAGAGAAATAAA 156

OY 481 GATGATGATGATGAA 495
    | | | | | | | | | |
Db 157 GATGATGACACTGAA 171
```

Search completed: March 5, 2003, 12:37:33
Job time : 173 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:29:14 ; Search time 21 seconds
(without alignments)
2554.429 Million cell updates/sec

Title: US-10-037-311a-1

Perfect score: 3004
Sequence: 1 MDONSRRSSPIRTTGGSGS.....GTLVPHVHCEDISWGLKLV 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3004	100.0	558	2 T02704	hypothetical prote
2	1706.5	56.8	526	2 A86274	F7A19.15 protein -
3	1679	55.9	539	2 T02705	hypothetical prote
4	1613.5	53.7	500	2 D86274	hypothetical prote
5	1583	52.7	537	2 B86274	F7A19.16 protein -
6	1528	50.9	503	2 D84528	hypothetical prote
7	1492	49.7	533	2 B84528	hypothetical prote
8	1411.5	47.0	440	2 B84527	hypothetical prote
9	1302	43.3	435	2 B86274	hypothetical prote
10	1124.5	37.4	562	2 A86773	hypothetical prote
11	112.5	3.7	2149	2 C96695	ribulose biphosph
12	110	3.7	529	2 T23431	hypothetical prote
13	110	3.7	658	2 A64584	hypothetical prote
14	110	3.7	1102	2 S35617	hypothetical prote
15	107	3.6	567	2 C82183	probable origin-bi
16	106.5	3.5	80	2 A84528	catalase VC1585 (1
17	106	3.5	3844	2 A84528	hypothetical prote
18	105	3.5	381	2 T18402	asparagine/asparta
19	104.5	3.5	1583	2 T14176	hypothetical prote
20	103.5	3.4	1209	2 T16663	probable phosphati
21	102	3.4	832	2 H72278	hypothetical prote
22	100.5	3.3	409	2 E86017	alpha-mannosidase-
23	100.5	3.3	494	2 E91171	probable beta-keto
24	97	3.2	594	2 B71893	excinuclease ABC c
25	97	3.2	678	2 C83878	methy1-accepting c
26	97	3.2	2044	2 A81180	probable peptidogl
27	97	3.2	4717	2 T41581	hypothetical colle
28	96	3.2	594	1 E64622	excinuclease ABC c
29	96	3.2	811	2 A80680	probable dimethyl

ALIGNMENTS

30	95.5	3.2	835	2 T05259	probable disease r
31	95	3.2	326	4 S61652	hypothetical prote
32	95	3.2	1802	2 G71616	hypothetical prote
33	94	3.1	717	2 T27066	hypothetical prote
34	94	3.1	1333	2 S30356	hypothetical prote
35	94	3.1	1376	1 J01534	CDC25 protein homo
36	93.5	3.1	808	2 F90915	E2 glycoprotein pr
37	93.5	3.1	808	2 C85764	probable oxidoredu
38	93.5	3.1	808	2 F64914	probable oxidoredu
39	93.5	3.1	1145	2 T33606	dimethylsulfoxide
40	93	3.1	455	2 S61159	hypothetical prote
41	93	3.1	673	1 VCPV85	CDC40 protein - ye
42	93	3.1	837	1 A56681	coat protein VPI -
43	93	3.1	1338	2 T18416	penicillin amidase
44	93	3.1	1418	2 S64918	hypothetical prote
45	93	3.1	1817	2 D71606	hypothetical prote

RESULT 1
T02704
hypothetical protein At2g03220 (imported) - Arabidopsis thaliana
N:Alternate names: hypothetical protein T18E12.11
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02704; G84445
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, September 1998
A:Description: Arabidopsis thaliana chromosome II BAC T18E12 genomic sequence.
A:Reference number: 214702
A:Accession: T02704
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-558 <ROU>
A:Cross-references: EMBL:AC005313; NID:g3548797; PID:g3548808
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-558 <STO>
A:Cross-references: GB:AE002093; NID:g3548808; PID:AC34480.1; GSPDB:GN00139
A:Genetics:
A:Gene: T18E12.11; At2g03220
A:Map position: 2
A:Introns: 93/1

Query Match
Best Local Similarity 100.0%; Score 3004; DB 2; Length 558;
Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDONSRRSSPIRTTGGSGSVNFSSELIQMYLSSGIMKLRFTTCTIVSVLAFSM 60
DB 1 MDONSRRSSPIRTTGGSGSVNFSSELIQMYLSSGIMKLRFTTCTIVSVLAFSM 60
QY 1FH0HPSDSNRIMGFAEAVVLDAGVFPNVTNINSOKLLGGLASGPDSCSRSQSVHY 120
DB 1FH0HPSDSNRIMGFAEAVVLDAGVFPNVTNINSOKLLGGLASGPDSCSRSQSVHY 120
QY 61 ITH0HPSDSNRIMGFAEAVVLDAGVFPNVTNINSOKLLGGLASGPDSCSRSQSVHY 120
DB 61 ITH0HPSDSNRIMGFAEAVVLDAGVFPNVTNINSOKLLGGLASGPDSCSRSQSVHY 120
QY 121 RRPSPYKPSYLLSKLRNKKRSGPTESYKALKOLOEHIIDGDECKYVWVWISFS 180
DB 121 RRPSPYKPSYLLSKLRNKKRSGPTESYKALKOLOEHIIDGDECKYVWVWISFS 180
QY 181 GICNRILSLASVFLVALLTDRLVLLVDRGKMDLDFCEPFLGSWLLPLDFPMTDQFDGIN 240
DB 181 GICNRILSLASVFLVALLTDRLVLLVDRGKMDLDFCEPFLGSWLLPLDFPMTDQFDGIN 240

RESULT 4
D86274
hypothetical protein F7A19.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: D86274
R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Hutzar, I.
Nature 408, 816-820, 2000
C.A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
C.A: IL, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86274
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-500 <STO>
A:Cross-references: GB:AE005172; NID:95080784; PIDN:AMD39294.1; GSPDB:GN00141
A:Genetics:
A:Map position: 1

RESULT 5
B86274
F7A19.16 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C:Accession: B86274
R:Theologis: A.; Ecker, J.R.; Palm, C.J.; Federpsiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
C.A: Authors: Salzb erg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallonker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:1130712
A:Accession: B86274
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-537 <STO>
A:Cross-references: GB:AE005172; NID:95080783; PIDN:AAD39293.1; GSPDB:GN00141
A:Genetics:
A:Map position: 1

Query Match	52.7%	Score 1583:	Db 2:	Length 537:
Best Local Similarity	59.1%	Pred. No. 3.7e-120:		
Matches 317:	Conservative 71:	Mismatches 122:	Indels 26:	Gaps
QY	10MKKLSSGTMKLTFRRTFTCLIVFSLVAFSMIFPHQHPDSNRIMGFARVILADGVFPN	88		
Db	17 LKMKLT-----LTIVFSGLTLWVSIVLSFS-----NDFNNOLLVATSNVSRESFTR	63		
QY	89 VTNINSDKLLGLLGSAGEDSDCSLSRY-QSVHYRRKPSYPSSYLLSKLRNYEKLHKRCG	147		
Db	64 -----DRLIGLTLNADDEGCSLSRYQOSSLYLRASPPKPSYELVSKLSRYEKLHKRCG	117		
QY	148 PGTSEYKKAKLQK--DQEIHIDSD--GECKYVYVWISFGSGGNLTLSAFYALALTDRLV	203		
Db	118 PGTDAKKKATELLGHDDENYVASKSVGECKRYIYVAVVYGGNRLITLTASVETALLTERVV	177		
QY	204 LVDRKQKDDLECEPFGLSGMWLPRLDPFPTDQFDGLNOBSSRCYGMVNOYIDTEGTL	263		
Db	178 LVDSQKDIISDLFCFEPFGTSMWLPRLPEFLPMKQIDGYNKYSRCYGTMLNQAINSTLPLP	237		
QY	264 HLVLHLVHDYGDHDMFCEGSDFTIGKVPWLIVKTDNVFVSLMLPGFDDELKLPFO	323		
Db	238 HLVLHILDSRNDKMFCEQKQSDLVDAKPMWLIVKANYFVPSLMLNPTFGFELMKLFPQ	297		
QY	324 KATVHNHLGRVLFHPTNOVWGLVTRYEARVLSHADEKIGIQVRFVEDRGPQOHVMDQIS	383		
Db	298 KEAVVHNHLARVLFHPTNOVWGLITRSYNAYLSRADETLGIDIRVFSDRAGVQOHVMDQV	357		
QY	384 SCTQKELLPEVDTLVFSRHNV-TPKHAVLVYSLMAGYAEMLKSMYETVTSGETIG	442		
Db	358 ACTRENELLPEPAOEEPRVNISSROKLAVALVYSLPYSETLKMNWERSSTGEIIE	417		
QY	443 VHPQSEGYOQOTEKRMHNGKALAEVYLLSLTDNLVYLSAMSTEGYVAAQGGGKLPMLVLP	502		
Db	418 VYQSGEYVQOTDKLHQQKALAEVYLLSLTDK IYTSASSTGYVAHSLGGKLPMLVLP	477		
QY	503 ENRTTPDPSCGRAMSEPCFHSPPFYDCKAKTGIDGTLVPHVRHCEDI-SWGLKL	557		
Db	478 TGPRTAPDPCIGDSTMDPCHLPPSHGCEPEWGTSGTSMVPEVLRCEBDNGNGLK	533		

Db 406 --QMSDVLNKKEMAAQOFVKS 425

RESULT 13

hypothetical protein HP0513 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: A64584

R:Tomb, J.F.; White, O.; Kellavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodde, A.; McKen-

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: A64584

A:Molecule type: DNA

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Residues: 1-658 <TOM>

A:Cross-references: GB:AE000565; GB:AE000511; NID:g2313616; PIDN:AA007579.1; PID:g231362

Query Match 3.7%; Score 110; DB 2; Length 658;

Best Local Similarity 17.8%; Pred. No. 1.2;

Matches 102; Conservative 79; Mismatches 191; Indels 202; Gaps 23;

37 GTMKLTFFETTCCLYFSVLVAFSMFIHQPSDSNRIMGFAERVLADAGVFPVNTINSDK 96

Db 28 GQORIT-TITLLLA-----LRNHLSEVEILKFSRKELSYLIN--SKRDK 74

Qy 97 LIGGLASGFEDSCLSRYSQSVHVRKPSRPSYSLIKLNVEKLRRCGPGTESYKKA 156

Db 75 KF-RILSESQKDTLSIDK-NKRKPS--EPSYIVENFELFEKWIENDKLETIRKG 130

Qy 157 LKQDQEHIDGDECKYVWIS-----FSGIGNRILSLASVFLYALLDRVLV 205

Db 131 LKLT-----MTVWISLDKGDQDPQLIFESMNSKDIETQT--DLIRNYIAME 175

Qy 206 DRGKMDLFCERFLGMSWMLRPLDEPMTDOF----- 236

Db 176 TEVEQGEFYQYKRAMEREQNETLNFNRVRIYLTIKIGIPNEKRVYEAFKDYRQK 235

Qy 237 ---DGLNQESSRCYGVW---KNQVIDTEGTLN-----LYLHVHDYGDH 276

Db 236 GIEIDILKQLOKYGCGYQIAFKKEDDKLNLKALSLVNLMDIYPLLEISDYKD- 294

Qy 277 DKMFCEBDQFFIGKVPMLVKTQNYFVPSLMLDPF-----DDELNKLFP--- 322

Db 295 -----GVLKQD--FIPITYLIESYICRAVCGIGTNSINKVFPSTK 335

Qy 323 --OKATVFHNLGRYLF-----HPTNQVMGLV-----TRYEAVLSHNDKIG 362

Db 336 HIQKDEYKSLKANHVCCLTERQRPNNDEFKKLFTTIDFYFKKNKYFLERLENEDTKEP 395

Qy 363 IQVRVDEDPGRFOHVMQDISCTQKEKLRPEVDLVERSRHV--NTPKRAVLVTSUNA 420

Db 396 V-----DTQKCNIEHIMQTLTLPWQBDLGENFOAIHEKYLHTIGN-----LTLTGYSN 444

Qy 421 GYAEV-----LKSM-----YVEYPTST 437

Db 445 KYSNNSFOEKRDMEKGFQSSSLKLNQSLKDLSEFGEKEIERRASDLADWALKIMYPILE 504

Qy 438 GEIGVHOPSOEQYQTEKKMKHNGKALAEWLLS 471

Db 505 AETLEEVYFKKKEKKKEEYKIKKKEKKVYDLS 538

RESULT 14

335617

probable origin-binding protein - African swine fever virus

C:Species: African swine fever virus, ASFV

C:Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 08-Oct-1999

C:Accession: S35617

R:Sussman, M.D.; Lu, Z.; Kutish, G.F.; Afonso, C.A.; Rock, D.L.

Nucleic Acids Res. 21, 2254, 1993

A:Title: The identification of an African swine fever gene with conserved hellicase mo

A:Reference number: S35617; MUID:93275762; PMID:8389041

A:Accession: S35617

A:Molecule type: DNA

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Residues: 1-1102 <SUS>

A:Cross-references: EMBL:L12174; NID:g289178; PIDN:AAA42719.1; PID:g289179

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1993

Query Match 3.7%; Score 110; DB 2; Length 1102;

Best Local Similarity 19.0%; Pred. No. 2.7;

Matches 103; Conservative 94; Mismatches 210; Indels 136; Gaps 26;

15 TTGGSKSVNFSLELD-MKYSSGTMKLTFRFTTCCLYFSVLVAF-----SMIFQNH 65

Db 251 TTSSSCDSIQOSELREVLF--AGTLCNHC--LRVYKNLVLEKRTSPSYCEICKRMH 305

Qy 66 PSDNRINGFAERVLADAGVFPVNTINISDKLLGULLA--SGFDE---DCLISR---YQ 116

Db 306 DKDNTLLI---RVTKGRKVVQCHRDHDKHSLMCSLSTNNFVETVEYQVMSKIEVHE 360

Qy 117 SVHYKPSRPSYSLIKLNVE-----KLHRCGPGTESYKKAQKQDQEHIDGPG 169

Db 361 SLTFEELPDQKHIDESSMREYERVPLLVKAQKIKGTIO---LRNYLQKTYGND 415

Qy 170 ECKY--VWISFGIGNRILS--LASVFLYALLT-----DRVLVDKQMDLFCERF 219

Db 416 ISKQOTIRFVFQRFQSKNIQRLPNFTLYSEVGTDLDSYERVIL-----QVESL----F 466

Qy 220 LGMSLLLPLDPMTPQDFGL-NQESSRCYGVW-----VKNOYIDTEGTLNHLV 266

Db 467 RLTSTAEPVLDLILDEVESIFNQFNSGLHKYFAPSAFEMMLETANVHICIDAMVIGRT 526

Qy 267 LHLVNDY-GD-----HDKMFCEGDDTFGKVPMLVKTQNDVFPVSLMLIP 311

Db 527 YNIIQRFQGVPIFFHNMQYQKQNDMYFTTSSREIIMNLKDLLEDKKIYIPNLSME 586

Qy 312 G--FDDELNKLFPQKATVFN-----HLGRYLFHPTNQVMGLV-----TRYEAVLSHAD 358

Db 587 ARLLQTFIQKKFPEKIKGFYSKSSAHERESHFNVSYYWGLIDILYITPTISAGVSD 646

Qy 359 EKIGIQVAVPEDEPGRFOHVMQDISCTQ-----KEKLRPEVDLV 399

Db 647 KR-----FDVLGFEFNMSCDVEPTCCOMLGRVRELKSKCYKICLOGKONYFPET---I 696

Qy 400 ERSRHVNTPKHKAVALVTSINAGYAENLKSMEWYPTSTGEIIGVHQPSOEGYQTEKKMH 459

Db 697 EDIEMFTLQKKDQLTQTLISN-----HQLSTFYCKEKGRTPIYTYTHLMLETRIOH 749

Qy 460 NGK 462

Db 750 LSK 752

RESULT 15

C82183

catalase VC1585 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: C82183

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

Charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: C82183

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-567 <HEI>

A:Cross-references: GB:AE004235, GB:AE003852, NID:99656083; PIND:AAF94739.1; GSPDB:GM001
A:Experimental source: serogroup O1, strain N16561, biotype El TOR
C:Genetics:
A:Gene: VC1585
A:Map position: 1
C:Superfamily: catalase

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:29:59 ; Search time 31 Seconds

(without alignments)
529.613 Million cell updates/sec

Title: US-10-037-311a-1

Perfect score: 3004
Sequence: 1 MDNSYRRSSPIRTTGGT.....GIVPHVHCEDISWGLKLV 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	3.1	723	4 US-09-134-001C-5060	Sequence 5060, Ap
2	92.5	3.1	993	4 US-08-836-667B-30	Sequence 30, Appl
3	92	3.1	809	4 US-09-186-276B-58	Sequence 58, Appl
4	92	3.1	809	4 US-08-842-445-58	Sequence 58, Appl
5	92	3.1	809	4 US-09-186-188B-58	Sequence 58, Appl
6	90	3.0	758	2 US-08-222-617A-6	Sequence 6, Appl
7	90	3.0	3666	2 US-08-222-617A-12	Sequence 12, Appl
8	90	3.0	3727	2 US-08-222-617A-27	Sequence 27, Appl
9	90	3.0	3727	2 US-08-222-617A-2	Sequence 2, Appl
10	89.5	3.0	1285	4 US-09-268-140-3	Sequence 3, Appl
11	89	3.0	1285	2 US-08-540-406-6	Sequence 6, Appl
12	89	3.0	1285	2 US-08-656-055-6	Sequence 6, Appl
13	89	3.0	1285	4 US-08-954-668-6	Sequence 6, Appl
14	89	3.0	1285	4 US-08-918-658-6	Sequence 6, Appl
15	89	3.0	1285	5 PCT-US95-13233-6	Sequence 6, Appl
16	88.5	2.9	515	3 US-08-369-822C-23	Sequence 23, Appl
17	88.5	2.9	515	3 US-08-582-776C-38	Sequence 38, Appl
18	88.5	2.9	515	3 US-08-434-831B-35	Sequence 35, Appl
19	86.5	2.9	771	4 US-09-462-284-2	Sequence 2, Appl
20	85	2.8	461	4 US-09-134-001C-5311	Sequence 5311, Ap
21	85	2.8	584	1 US-08-179-738-7	Sequence 7, Appl
22	85	2.8	584	1 US-08-628-145-7	Sequence 7, Appl
23	85	2.8	591	1 US-08-179-738-5	Sequence 5, Appl
24	85	2.8	591	1 US-08-628-145-5	Sequence 5, Appl
25	85	2.8	596	1 US-08-179-738-2	Sequence 2, Appl
26	85	2.8	596	1 US-08-628-145-2	Sequence 2, Appl
27	85	2.8	777	2 US-08-874-678-3	Sequence 3, Appl

28	85	2.8	777	3 US-08-643-839-3	Sequence 3, Appl
29	85	2.8	777	4 US-09-348-886-3	Sequence 3, Appl
30	85	2.8	1298	1 US-08-222-616-33	Sequence 33, Appl
31	85	2.8	1298	1 US-08-340-011-2	Sequence 2, Appl
32	85	2.8	1298	4 US-08-901-710-2	Sequence 2, Appl
33	85	2.8	1298	4 US-08-446-648-33	Sequence 33, Appl
34	85	2.8	1298	5 PCT-US95-04228-33	Sequence 33, Appl
35	85	2.8	1363	1 US-08-340-011-4	Sequence 4, Appl
36	85	2.8	1363	2 US-08-874-678-32	Sequence 32, Appl
37	85	2.8	1363	3 US-08-643-839-32	Sequence 32, Appl
38	85	2.8	1363	3 US-08-901-710-4	Sequence 4, Appl
39	85	2.8	1363	4 US-09-348-886-32	Sequence 32, Appl
40	84.5	2.8	848	1 US-08-045-806-4	Sequence 4, Appl
41	84.5	2.8	848	1 US-08-366-051B-4	Sequence 4, Appl
42	84	2.8	591	1 US-08-179-738-10	Sequence 10, Appl
43	84	2.8	591	2 US-08-628-145-10	Sequence 10, Appl
44	84	2.8	595	1 US-08-171-718-16	Sequence 16, Appl
45	84	2.8	595	3 US-08-478-087-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1					
US-09-134-001C-5060					
; Sequence 5060, Application US/09134001C					
; Patent No. 6380370					
; GENERAL INFORMATION:					
; APPLICANT: Lynn Doucette-Stamm et al					
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC					
; FILE REFERENCE: GTC-007					
; CURRENT APPLICATION NUMBER: US/09/134,001C					
; PRIOR APPLICATION NUMBER: 1998-08-13					
; PRIOR FILING DATE: 1997-11-08					
; PRIOR APPLICATION NUMBER: US 60/064,964					
; PRIOR FILING DATE: 1997-08-14					
; NUMBER OF SEQ ID NOS: 5674					
; SEQ ID NO 5060					
; LENGTH: 723					
; TYPE: PRT					
; ORGANISM: Staphylococcus epidermidis					
US-09-134-001C-5060					
Query Match 3.1%; Score 93; DB 4; Length 723;					
Best Local Similarity 20.9%; Pred. No. 1.1;					
Matches 110; Conservative 60; Mismatches 168; Indels 168; Gaps 28;					
QY	113	SRYSVHYRRKPPSPSSYLIS----	KLBNYEKLKRCGPTESYKKALKOLDOE--	HTD	166
DB	212	SHODEINTFKPKRYTTLSTINVDGYDKL--	NOOKRYKD-----	KKLELLEHETKHOE	262
QY	167	GDGECKYVWISFSGIGRNILSLASVFLYALTD	RVLDVRGKMDDLFCPEFLGMSWL	L	226
DB	263	G-----	-----	-----	-----
QY	227	PLDPEMTDQFDGLNDESSKCYWYKNOYITDE	GLSHLY--LHLYVDGDHMKFCEGD		285
DB	279	PL-FMLTD-----	LOOEAVKRYKMPK-----	ETLNLLOHLYERHKLVTYPRTSNYLTDOM	329
QY	286	QYFKKVPMLYKTYNTYFVPSLMLPGFDELNK	LFPKATVF-----	HH-----	330
DB	330	VDTDERLRAIATD-----	-----	-----	-----
QY	331	-----	-----	-----	-----
DB	382	PSIEDQSREFKTYMLIERPLENLMNPYLEVLT	HAQLKDYHFLVKKIKQL-----		436
QY	371	DGCPQHWWDIISCTQ-----	KEKLLPEYDVLVENSRYHNPFKH--	KAVLYTSL-----	418
DB	437	---GYKALKDOLSHTLTHSFKEGOLFVKVRIE	HNHETKAPDEYFNESGLKAMENPONH		493

```

OY 419 ---NAGGAENIKSMWYEPTSTGIIIV-----HQPQOEYOQTEKKMHNGKA 463
    119 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 494 IDLNKRKYATKLFK-----HTGIGIVATRADILIEKLFPMNMLESDRGIKYTSNGQ 545
    119 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 464 LAEWLITLDTNLTYSAM-STFGVYAGGLGSLKWT-----LYRPE 503
    119 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 546 ILELSPSELSTPILTAQHEEKLMLIEKKYNSQFETDMKNFTFKYVNRKIKSSEQKYRD 605
    119 111 111 111 111 111 111 111 111 111 111 111 111 111
OY 504 NRTTFD-PSGCRAM-----SMEPCSHSPPEYDCAKGTGIDTGT 540
    119 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 606 NLTTTECTCGKFMIKYTKNGQLVC-QDD--KCKTKNRIOKRT 647
    119 111 111 111 111 111 111 111 111 111 111 111 111 111

```

```

US-08-836-687B-30
: RESULT 2
: Sequence 30, Application US/08836687B
: Patent No. 6448034
: GENERAL INFORMATION:
: APPLICANT: Gasson, Michael John
: APPLICANT: Dodd, Helen Mair
: TITLE OF INVENTION: PRODUCTION OF VARIANT NISIN
: FILE REFERENCE: 20747/70
: CURRENT APPLICATION NUMBER: US/08/836,687B
: CURRENT FILING DATE: 1995-11-20
: NUMBER OF SEQ ID NOS: 51
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 30
: LENGTH: 993
: TYPE: PRT
: ORGANISM: Lactococcus sp.
US-08-836-687B-30

```

Query Match	3.1%	Score	92.5	DB	4	Length	993
Best Local Similarity	19.1%	Pred. No.	2.1				
Matches	90	Conservative	63	Mismatches	152	Indels	167
						Gaps	23

```

QY      89  VTNIINSDKLLIGGLASGDEDSCLSRQSYHYKR--SPKPSYSLLSKIRANERKHKKC 146
Db      224  ISNLOKD-----LLSDFSMNTFLTKVEALDEDEKKYIIPLKVKQKFOESESIE-----I 272
QY      147  GPGFESYKALKOLDQ-----EHIDGDECKYUWWTISFSGNRIILSIAS 191
Db      273  GEGIEKLEKEIOEMSQILENDNYIQLDILSDSEINFDFVKQOOLEHIAELFNTKTSVAR 332
QY      192  VFL--YALLDPRVLL---VDGRKMMDLFCBPFLGMSWILLPDPFMTDOEDG-----LN 240
Db      333  TYLDLY---KDKFLEKGVDEVOYLELPFSTP-GIGAPARYNHPNDYSESEPSITLYAS 388
QY      241  QESSRCYGYM---VKN-QVIDTEGTLSH-----LYLHVIHGYGDHDK 279
Db      389  EEREKELYSMYEAUVKNNHNYIINLDDESHQOKMDLEKSSLOGLBELFLINAKEX---END 445
QY      280  FFCGDDOTFLICKRWMLVYKIDNTFVPSLMLIPGFDELNKLFPQKATVFHNHIGRYLHPT 339
Db      446  IFLIGD-----IVGNNN-----LGGASGRFSALISPE----- 471
QY      340  NOVWGLVTRYVEAVLISHADEKIGIOYRVFDEDPGFQHWADIOSCTQOKEXLLPEVDLY 399
Db      472  -----LTSVHRTIYDSEVER-----ENKEKEITSC-----ELVFLP 501
QY      400  ERSNHN-----TPKKAVLYTSLNAGYAENLKSWMYEPYTGIIIGV 443
Db      502  ENIHAWMHTSILMRKVLPEFTSTSHNEVLNINIIYIDE-KEKFVARDISTQEVLEKF 559
QY      444  HQPQEQEQOOTEKKMHN-----GKALAEMLYLLSLTDLYLSAMS-----TGYV 487
Db      560  YITS-----MYKTLFSLNELRFLLEISIDDKFGMLPELVIYRDYDI 601

```

RESULT 3
US-09-186-276B-58
; Sequence 58, Application US/09186276B
; Patent No. 6388173

```

1  / GENERAL INFORMATION:
2  / APPLICANT: Benfey, Philip
3  / APPLICANT: Dilaurenzio, Laura
4  / APPLICANT: Wyszocka-Diller, Joanna
5  / APPLICANT: Malamy, Jocelyn E.
6  / APPLICANT: Pysh, Leonard
7  / APPLICANT: Helariutta, Tarmo
8  / TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
9  / FILE REFERENCE: 5914-075-999
10 / CURRENT APPLICATION NUMBER: US/09/186,276B
11 / PRIOR APPLICATION NUMBER: 08/842,445
12 / PRIOR FILING DATE: 1997-04-24
13 / PRIOR APPLICATION NUMBER: 08/638,617
14 / PRIOR FILING DATE: 1996-04-26
15 / NUMBER OF SEQ ID NOS: 79
16 / SOFTWARE: PastSeq for Windows Version 3.0
17 / SEQ ID NO 58
18 / LENGTH: 809
19 / TYPE: PRT
20 / ORGANISM: Arabidopsis thaliana
21 / FEATURE:
22 / NAME/KEY: VARIANT
23 / LOCATION: (1)...(809)
24 / OTHER INFORMATION: Xaa = Any Amino Acid
25 / US-09-186-276B-58

```

[illegible]

```

RESULT 4
US-08-842-445-58
; Sequence 58, Application US/08842445A
; Patent No. 6441270
; GENERAL INFORMATION:
; APPLICANT: Benfey et al.
; TITLE OF INVENTION: Sarcosine Gene, Promoter and Uses
; TITLE OF INVENTION: Theatof
; FILE REFERENCE: 5914-056-999
; CURRENT APPLICATION NUMBER: US/08/842,445A
; CURRENT FILING DATE: 1997-04-24
; EARLIER APPLICATION NUMBER: 08/638, 617
; EARLIER FILING DATE: 1996-04-26

```

NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 58
LENGTH: 809
TYPE: PRF
ORGANISM: Plant
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(809)
OTHER INFORMATION: xaa = Any Amino Acid
US-08-842-445-58

Query Match 3.1%; Score 92; DB 4; Length 809;
Best Local Similarity 21.7%; Pred. No. 1.7;
Matches 84; Conservative 52; Mismatches 179; Indels 72; Gaps 18;

QY 66 PSDSNRIMGFAEARVLD-----AGVFP-----NTNINSKLLGGLASGDE-----108
DB 63 PPSSNQTLGLANGFYLDLDFSSLDPEAYPSONNNNNINNKAAVDLLSSSSDDADFS 122
QY 109 DSCLSRQSVHYRRSPYKPSSTYLS-KLRNYEK-LHKRCGPGESEYKKALKQDOEHI- 165
DB 123 DSVLKYSQVLMEDMEKPCMFHDALALQAEKSLTEALGEKDPSSSSASVDHPERLA 182
QY 166 --DGDGCKYVWVWISFSGLGNRLSLASVFLYALITDVRLLVDRGKMDDLFCPEFLGMS 223
DB 183 SHSPDGSCS-----GGAFSDYAS-----TTTTSSDSHWSVDGLENRP-----S 221
QY 224 WL---LPDLPMTDQFGLNOESSRCYGYWVKNOVIDTEGTLSHLYLHLVHDYGDHDKMF 280
DB 222 WLHTPMSNF---VFQSTRSNSVTGGGGGNSAVYSGSGDDLVSNMFKD--DELAHQ 275
QY 281 FCEGDO---FPIGVPMILVKTQNYFVPSLMLIPGFDDELNKLFP--PQKATVFHHLGRY 334
DB 276 FKKGVEASRFLPKRSOLFIDVDST-IP---MNSGSKENGSEVVKTEKKDETEHHHHS 331
QY 335 LFHPTNQWGLVTRYEAYLSHADEKIGIQRVDEDEPGFQHYMDQISSCTQKELPE 394
DB 332 YAPPPNLTGKKSHMRDEDEDFEERSNKOSAVYEE-SELSEMFDMNFILOGPKPVC-- 388
QY 395 VDTLVERSRYVTPKHKAVLVTSLNAG 421
DB 389 -----ILNQNFPTESAKVYTAQSNQ 408

RESULT 5
US-09-186-188B-58
Sequence 58, Application US/09186188B
Patent No. 6455672
GENERAL INFORMATION:
APPLICANT: Benfey et al.
TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
FILE REFERENCE: 5914-074-999
CURRENT APPLICATION NUMBER: US/09/186,188B
CURRENT FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 08/842,445
PRIOR FILING DATE: 1997-04-24
PRIOR APPLICATION NUMBER: 08/638,617
PRIOR FILING DATE: 1996-04-26
NUMBER OF SEQ ID NOS: 79
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 58
LENGTH: 809
TYPE: PRF
ORGANISM: Plant
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(809)
OTHER INFORMATION: xaa = Any Amino Acid
US-09-186-188B-58

Query Match 3.1%; Score 92; DB 4; Length 809;

Best Local Similarity 21.7%; Pred. No. 1.7;
Matches 84; Conservative 52; Mismatches 179; Indels 72; Gaps 18;

QY 66 PSDSNRIMGFAEARVLD-----AGVFP-----NTNINSKLLGGLASGDE-----108
DB 63 PPSSNQTLGLANGFYLDLDFSSLDPEAYPSONNNNNINNKAAVDLLSSSSDDADFS 122
QY 109 DSCLSRQSVHYRRSPYKPSSTYLS-KLRNYEK-LHKRCGPGESEYKKALKQDOEHI- 165
DB 123 DSVLKYSQVLMEDMEKPCMFHDALALQAEKSLTEALGEKDPSSSSASVDHPERLA 182
QY 166 --DGDGCKYVWVWISFSGLGNRLSLASVFLYALITDVRLLVDRGKMDDLFCPEFLGMS 223
DB 183 SHSPDGSCS-----GGAFSDYAS-----TTTTSSDSHWSVDGLENRP-----S 221
QY 224 WL---LPDLPMTDQFGLNOESSRCYGYWVKNOVIDTEGTLSHLYLHLVHDYGDHDKMF 280
DB 222 WLHTPMSNF---VFQSTRSNSVTGGGGGNSAVYSGSGDDLVSNMFKD--DELAHQ 275
QY 281 FCEGDO---FPIGVPMILVKTQNYFVPSLMLIPGFDDELNKLFP--PQKATVFHHLGRY 334
DB 276 FKKGVEASRFLPKRSOLFIDVDST-IP---MNSGSKENGSEVVKTEKKDETEHHHHS 331
QY 335 LFHPTNQWGLVTRYEAYLSHADEKIGIQRVDEDEPGFQHYMDQISSCTQKELPE 394
DB 332 YAPPPNLTGKKSHMRDEDEDFEERSNKOSAVYEE-SELSEMFDMNFILOGPKPVC-- 388
QY 395 VDTLVERSRYVTPKHKAVLVTSLNAG 421
DB 389 -----ILNQNFPTESAKVYTAQSNQ 408

RESULT 6
US-08-222-617A-6
Sequence 6, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-Apr-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 758 amino acids
TYPE: amino acid
STRANDEDNESS: single

```

1  TOPOLOGY: linear
2  MOLECULE TYPE: protein
3  FRAGMENT TYPE: Internal
4  ORIGINAL SOURCE:
5  ORGANISM: Penicillium chrysogenum
6  FEATURE:
7  NAME/KEY: Domain
8  LOCATION: 1..758
9  OTHER INFORMATION: /label=Domain II
10 OTHER INFORMATION: /note="Domain II of ACV Synthetase from
11 OTHER INFORMATION: Penicillium chrysogenum; aa 1397-2154"
12 US-08-222-617A-6
13
14 Query Match 3.08; Score 90; DB 2; Length 758;
15 Best Local Similarity 20.78; Pred. No. 2.5;
16 Matches 67; Conservative 35; Mismatches 99; Indels 122; Gaps
17
18 QY 71 RIMGAEARVLDAGFPWPVTNINSDKLLGLGASGPDSDLSRGSVHYRKPSPKPS-129
19 | | | | | : : : | : | | | | |
20 DB 132 RIKGMAASGTL---LPSVLPAND-----SKWS---VSNPSLSRST 168
21 | | | | | : : : | : | | | | |
22 QY 130 --SYLSIKLRNYEKLHKRCGPSTESYKALKQLDQEHIDGGECKYVWVTSFSGLGNRL 187
23 | | | | | : : : | : | | | | |
24 DB 169 DLAVIIV-----SGTGPRKG-----VTVEHHGVNVLLQV 198
25 | | | | | : : : | : | | | | |
26 QY 188 SLAVFLYALLTDRLVLDNRKDMD--DLFCFFPLMSWLLPLDFMTDQFGLNDSS 244
27 | | | | | : : : | : | | | | |
28 DB 199 SLSKVEFLRDTDEVEILSFSSVYVDHVEQMTDAILNGQTLVLN-----DGMRDKE 251
29 | | | | | : : : | : | | | | |
30 QY 245 RCGYGMVNOYIDNEGTLSHLXLVLDYGDHDMFCEDG-----DTFTGKVP 293
31 | | | | | : : : | : | | | | |
32 DB 252 RLRYIEKNRRTYISGTPSVSMTEFSRFDHLRLRVDCVGSANSEVPDKIRTFHGLV- 310
33 | | | | | : : : | : | | | | |
34 QY 294 WLIVKTDNYEVPISMLIPGFDELNKLFPKATVFNHLGRYLFHPTNQVGLVTRYEAV 353
35 | | | | | : : : | : | | | | |
36 DB 311 -----INGGPTPEVSTTTKRLYLPF-PEKRM----- 335
37 | | | | | : : : | : | | | | |
38 QY 354 LSHADEKIGIOVR-----VPEDD 371
39 | | | | | : : : | : | | | | |
40 DB 336 ---DKSIGQGVHNSYLVNED 354
41
42 RESULT 7
43 US-08-222-617A-12
44 Sequence 12, Application US/08222617A
45 Patent No. 5882879
46 GENERAL INFORMATION:
47 APPLICANT: Veenstra, Annemarie E.
48 APPLICANT: Martin, Juan F.
49 APPLICANT: Garcia, Bruno D.
50 APPLICANT: Gutierrez, Santilago
51 APPLICANT: Barredo, Jose L.
52 APPLICANT: Von Doehren, Hans
53 APPLICANT: Palissa, Harriet
54 APPLICANT: Van Liempt, Henk
55 APPLICANT: Montenegro, Eduardo P.
56 TITLE OF INVENTION: A Method for Influencing Beta-Lactam
57 TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
58 TITLE OF INVENTION: Quantities of ACV Synthetase
59 NUMBER OF SEQUENCES: 27
60 CORRESPONDENCE ADDRESS:
61 ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
62 STREET: 300 South Wacker Drive
63 CITY: Chicago
64 STATE: Illinois
65 COUNTRY: USA
66 ZIP: 60606
67 COMPUTER READABLE FORM:
68 MEDIUM TYPE: Floppy disk
69 COMPUTER: IBM PC compatible
70 OPERATING SYSTEM: PC-DOS/MS-DOS
71 SOFTWARE: PatentIn Release #1.0, Version #1.25
72 CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/222,617A
; FILING DATE: 04-APR--1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97,157
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; HYPOCHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Penicillium chrysogenum
; FEATURE:
; NAME/KEY: protein /label= region
; LOCATION: 1..3666
; OTHER INFORMATION:
; OTHER INFORMATION: /note= "Region of ACV Synthetase; aa 62-3727"
US-08-222-617A-12

Query Match          3.0%; Score 90; DB 2; Length 3666;
Best Local Similarity 20.7%; Pred. No. 33;
Matches    67; Conservative   35; Mismatches   99; Indels 122; Gaps 13;

OY      71 RIMGAEARVLDAAGFPNVTINSKLLGLLASGFDESDCSRYSQVHYRKPSPKPS-129
           ||| | :|| :||| | ||| | ||| | ||| | ||| | ||| |
DB     1467 RKGMMAASGTL---LYPVLPRNP-----SKMS---VSNSPLSNST 1503
OY     130 --SYLSIKLRNYEKLRKRCGPCTESYKKALKQLDDEHIDGDGECKYYVVWISFGIGNRIL 187
           ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
DB     1504 DLAIIIT-----SGTGRPKG-----TVENHGIVNLQV 1533
OY     188 SLASFELVALLDRLYLVDRCRDMD---DLPCPEFLGSMLRLDFRPMTDPFDGLNQESS 244
           ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
DB     1534 SLSKFEGLRDDDEVILEFSNVFPHVEQMPTDALINGOTLLVIN---DGMRGKE 1586
OY     245 RCYGVMKNQVIDTEGTSLHLVLVHDYGDHKMFCEGD-----QTFIGKVP 293
           ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
DB     1587 RLRYRIEKNRTVYLSGTSSVSMYEFSREKHLLRRYDCVGAFSPVRDKIRETHGLV- 1645
OY     294 WLIVKTDNFYEPSLWLIPGFDELKLLEPPOKATVFHHLGRLFHPTNOVWGILTRYRAY 353
           ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
DB     1646 -----INGCPFEVSIYTHKRKYPF-BERRM----- 1670
OY     354 ISHADEKICIQVR----VFDED 371
           ||| | ||| | ||| | ||| | ||| | ||| | ||| |
DB     1671 ---DKSIGQQVHNSTSYLVLED 1689

RESULT 8
US-08-222-617A-27
; Sequence 27, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan P.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doefern, Hans
; APPLICANT: Palisac, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A method for influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
```

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3727 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-222-617A-27

Query Match 3.0%; Score 90; DB 2; Length 3727;
Best Local Similarity 20.7%; Pred. No. 34;
Matches 67; Conservative 35; Mismatches 99; Indels 122; Gaps 13;

OY 71 RIMGFAEARVLDAGVFPVNTNINSKLLAGLLASGFDESDCLSRQSVHYRRPSYKPS- 129
DB 1528 RIKGMAASGTL---LYPSVLPAHPD-----SKWS---VSNPSLSRST 1564
OY 130 --SYLISKLRNVEKLRKRCGPTESEYKRALQDQEHIDGDECKYVWVMSFGLGNRL 187
DB 1565 DLAVIITV-----SCTTGRPKG-----YVVEHNGVNNLOV 1594
OY 188 SIASVFLALLTDRLVLDVDRGKMD--DLCEPFLGMSWLLPLDFPMTDOFGDNGESS 244
DB 1595 SLSKVFGLMDTDEVLILFSNVVDFHFVEQMTDALINGQTLVLN-----DGMKGDE 1647
OY 245 RCYGVWVKNQVYDTEGTLSHLYLHLVHDYGDHDKMFECEGD-----QTFLGKVP 293
DB 1648 RLYRIEKRKYVLSGTSVSVSMYEFSPKRDHLRVDCVGEAFSESPVDKIRETHGLV- 1706
OY 294 WLIVKTDVNFVPSLMLIPGFDELNKLFPQKATVFHNLGRYLFHPTNQVGLVTRYEAY 353
DB 1707 -----INGYGPTEVSITTHKRLYFF-PERRR----- 1731
OY 354 LSHADEKIGIOVR-----VFDED 371
DB 1732 ----DKSIGQOVHNSTSYLNED 1750

RESULT 9
US-08-222-617A-2
Sequence 2, Application US/08222617A
Patent No. 5882879

GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3778 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-222-617A-2

Query Match 3.0%; Score 90; DB 2; Length 3778;
Best Local Similarity 20.7%; Pred. No. 35;
Matches 67; Conservative 35; Mismatches 99; Indels 122; Gaps 13;

OY 71 RIMGFAEARVLDAGVFPVNTNINSKLLAGLLASGFDESDCLSRQSVHYRRPSYKPS- 129
DB 1528 RIKGMAASGTL---LYPSVLPAHPD-----SKWS---VSNPSLSRST 1564
OY 130 --SYLISKLRNVEKLRKRCGPTESEYKRALQDQEHIDGDECKYVWVMSFGLGNRL 187
DB 1565 DLAVIITV-----SCTTGRPKG-----YVVEHNGVNNLOV 1594
OY 188 SIASVFLALLTDRLVLDVDRGKMD--DLCEPFLGMSWLLPLDFPMTDOFGDNGESS 244
DB 1595 SLSKVFGLMDTDEVLILFSNVVDFHFVEQMTDALINGQTLVLN-----DGMKGDE 1647
OY 245 RCYGVWVKNQVYDTEGTLSHLYLHLVHDYGDHDKMFECEGD-----QTFLGKVP 293
DB 1648 RLYRIEKRKYVLSGTSVSVSMYEFSPKRDHLRVDCVGEAFSESPVDKIRETHGLV- 1706
OY 294 WLIVKTDVNFVPSLMLIPGFDELNKLFPQKATVFHNLGRYLFHPTNQVGLVTRYEAY 353
DB 1707 -----INGYGPTEVSITTHKRLYFF-PERRR----- 1731
OY 354 LSHADEKIGIOVR-----VFDED 371
DB 1732 ----DKSIGQOVHNSTSYLNED 1750

RESULT 10
US-09-268-140-3
Sequence 3, Application US/09268140
Patent No. 6268176

GENERAL INFORMATION:
APPLICANT: Gemmill, Robert M.
APPLICANT: Drabkin, Harry A.
TITLE OF INVENTION: TRC8, A GENE RELATED TO THE HEDGEHOG RECEPTOR, PATCHED
FILE REFERENCE: 93445-00004
CURRENT APPLICATION NUMBER: US/09/268,140
CURRENT FILING DATE: 2000-03-12
PRIOR APPLICATION NUMBER: US 60/077,723
PRIOR FILING DATE: 1998-03-12
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 1286
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-268-140-3

Query Match 3.0%; Score 89.5; DB 4; Length 1286;
Best Local Similarity 20.6%; Pred. No. 6.8;
Matches 64; Conservative 41; Mismatches 81; Indels 125; Gaps 17;

```
QY 218 PRLGMSW-----LPLDPEPMDQFDGLN-----QESSRCGYM 250
    ||| ||| : : : ||| : : : |||
DB 672 PELMSWVKFELTVMGFLAALLSSLYASTRLQDGLDIDLVPKDSNEHFKFLDAQTRLFGEY 731
QY 251 VKNQYIDEGTSLH-LYLHLVHDYGDHDKMFCESDQTFIGVPMILVKTDMYFVPSLML 309
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 732 SMAYV--TQGNFEYPTQOQLLDY--HD-----SFV-RVPH-VIKNDNGGLPDEFWL 776
QY 310 IEGFDELNKLFPQKATVFNHIGRYLFHPTNQWGLVTRYEAYLSHADEKIGIQVRVD 369
    : : : ||| : : : ||| : : : ||| : : : |||
DB 777 L-----LF-----SEWLGNIQKIFD 791
QY 370 EDGPGQVHMDQISSCTQKEKLLPEVDT-----LVSRSHVNTPKHKAVALYS--LNA 420
    : : : ||| : : : ||| : : : ||| : : : |||
DB 792 E-----EYRDGRLTKECFPNASSDAILAYKLIQVTHVDNPDVKEVLVTRNLVNS 842
QY 421 GVAENLKSMY-----WEYPTSTGEITGVHQPQSEGVOQTEKKHNGKAL-----A 465
    : : : ||| : : : ||| : : : ||| : : : |||
DB 843 DGTINQRAFYNTLSAMATNDVFAYGASQGL--YEPKROYFHQPNEDLKIPKSLPLVYA 900
QY 466 EM--YLLSLTD 474
    : : ||| |||
DB 901 QMPFYLHGLTD 911
```

RESULT 11

US-08-540-406-6
; Sequence 6, Application US/08540406
; Patent No. 5837538

GENERAL INFORMATION:
APPLICANT: SCOTT, MATHEW P
APPLICANT: GOODRICH, LISA V
APPLICANT: JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/540,406
FILING DATE: 06-OCT-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-540-406-6

Query Match 3.0%; Score 89; DB 2; Length 1285;
Best Local Similarity 20.3%; Pred. No. 7.6;

Matches 63; Conservative 39; Mismatches 83; Indels 126; Gaps 16;

```
QY 218 PRLGMSW-----LPLDPEPMDQFDGLN-----QESSRCGYM 250
    ||| ||| : : : ||| : : : |||
DB 672 PELMSWVKFELTVMGFLAALLSSLYASTRLQDGLDIDLVPKDSNEHFKFLDAQTRLFGEY 731
QY 251 VKNQYIDEGTSLH-LYLHLVHDYGDHDKMFCESDQTFIGVPMILVKTDMYFVPSLML 309
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 732 SMAYV--TQGNFEYPTQOQLLDY--HDSF-----RVPH-VIKNDNGGLPDEFWL 775
QY 310 IEGFDELNKLFPQKATVFNHIGRYLFHPTNQWGLVTRYEAYLSHADEKIGIQVRVD 369
    : : : ||| : : : ||| : : : ||| : : : |||
DB 776 L-----LF-----SEWLGNIQKIFD 790
QY 370 EDGPGQVHMDQISSCTQKEKLLPEVDT-----LVSRSHVNTPKHKAVALYS--LNA 420
    : : : ||| : : : ||| : : : ||| : : : |||
DB 791 E-----EYRDGRLTKECFPNASSDAILAYKLIQVTHVDNPDVKEVLVTRNLVNS 841
QY 421 GVAENLKSMY-----WEYPTSTGEITGVHQPQSEGVOQTEKKHNGKAL-----A 465
    : : : ||| : : : ||| : : : ||| : : : |||
DB 842 DGTINQRAFYNTLSAMATNDVFAYGASQGL--YEPKROYFHQPNEDLKIPKSLPLVYA 899
QY 466 EM--YLLSLTD 474
    : : ||| |||
DB 900 QMPFYLHGLTD 910
```

RESULT 12

US-08-656-055-6
; Sequence 6, Application US/08656055
; Patent No. 6027882

GENERAL INFORMATION:
APPLICANT: SCOTT, MATHEW P
APPLICANT: GOODRICH, LISA V
APPLICANT: JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/656,055
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/540,406
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-656-055-6

Query Match 3.0%; Score 89; DB 3; Length 1285;

Query Match 3.0%; Score 89; DB 4; Length 1285;
Best Local Similarity 20.3%; Pred No. 7.6; Mismatches 83; Indels 126; Gaps 16;
Matches 63; Conservative 39;

QY 218 PFLGMSW-----LLPLDFPMTDQFDGLN-----QESSRCYGM 250
DB 672 PFLMSWVKFLTWGFLAALLSSLYASTRLDGLDITIDLVPKDSNEHKFLDAQTRLFGEFY 731
QY 251 VKNQVIDTEGTLSH-LYLHLVHDYGDHDKMFCEGDQTFIGKVPWLIVKTDNYEVPSTLML 309
DB 732 SMYAV--TQGNFEYPTQOQLRDY--HDSF-----RVPH-VIKNDNGGLPDPWL 775
QY 310 IPGFDELNKLFPQKATVFNHILGRYLFHPTNQWGLVTRYEAYLSHADEKIGIOVRVD 369
DB 776 L-----LF-----SEWLGNIQKIFD 790
QY 370 EDPGFQHVMDQISSCTQKELKLEVD-----LVRSRHVNTPKHKAVALTS--LNA 420
DB 791 E-----EYRDGRLTRKCEWFPNASSDAILAYKLIQVGHVNDVPDKELVITNRLVNS 841
QY 421 GYAENLKSMY-----WEYPTSTGEIIGVHQPDSQGYOQTEKKMHNGKAL---A 465
DB 842 DGLINQRAFYNYLSAMATNDVFAYGASQGL--YEPHQYFHQPNEDYDKIPKSLPLVYA 899
QY 466 EM--YLLSLTD 474
DB 900 QMPFYLHGLTD 910

RESULT 15

PCT-US95-13233-6
Sequence 6, Application PC/TUS9513233
GENERAL INFORMATION:
APPLICANT: THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY
TITLE OF INVENTION: Patched Genes and their use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohnach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13233
FILING DATE: 06-OCT-1990
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-13233-6

Query Match 3.0%; Score 89; DB 5; Length 1285;
Best Local Similarity 20.3%; Pred No. 7.6; Mismatches 83; Indels 126; Gaps 16;
Matches 63; Conservative 39;

QY 218 PFLGMSW-----LLPLDFPMTDQFDGLN-----QESSRCYGM 250
DB 672 PFLMSWVKFLTWGFLAALLSSLYASTRLDGLDITIDLVPKDSNEHKFLDAQTRLFGEFY 731
QY 251 VKNQVIDTEGTLSH-LYLHLVHDYGDHDKMFCEGDQTFIGKVPWLIVKTDNYEVPSTLML 309
DB 732 SMYAV--TQGNFEYPTQOQLRDY--HDSF-----RVPH-VIKNDNGGLPDPWL 775
QY 310 IPGFDELNKLFPQKATVFNHILGRYLFHPTNQWGLVTRYEAYLSHADEKIGIOVRVD 369
DB 776 L-----LF-----SEWLGNIQKIFD 790
QY 370 EDPGFQHVMDQISSCTQKELKLEVD-----LVRSRHVNTPKHKAVALTS--LNA 420
DB 791 E-----EYRDGRLTRKCEWFPNASSDAILAYKLIQVGHVNDVPDKELVITNRLVNS 841
QY 421 GYAENLKSMY-----WEYPTSTGEIIGVHQPDSQGYOQTEKKMHNGKAL---A 465
DB 842 DGLINQRAFYNYLSAMATNDVFAYGASQGL--YEPHQYFHQPNEDYDKIPKSLPLVYA 899
QY 466 EM--YLLSLTD 474
DB 900 QMPFYLHGLTD 910

Search completed: February 27, 2003, 15:32:58
Job time : 41 secs

Db 102 MEAQMFAFAVLAHLNHDYDDDKMFECDDQR-----LVNKRTDYYIVPSLIVTTFOD 154
 QY 316 ELNKLFPKATVFNHGLRYLPHPTN 340
 ||:||||: |||:|||||
 Db 155 ELDALEPBGAVFHYIGRLFFQAN 179

RESULT 2

Q98AU6 PRELIMINARY; PRT; 329 AA.
 AC 098AU6:
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
 GN Nodulation protein, NodZ.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
 Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003007; BAB5226.1; -
 KW Complete proteome.
 SQ SEQUENCE 329 AA; 37449 MW; 2460F90207D38287 CRC64;

Query Match 3.9%; Score 116.5; DB 16; Length 329;
 Best Local Similarity 21.1%; Pred. No. 0.11;
 Matches 78; Conservative 44; Mismatches 123; Indels 125; Gaps 15;

QY 165 IDGDECKYVWISFSGNRLSLASVFLALTLDRVLVD-RGK-DMDDLFCPEPLGM 222
 Db 3 VDRSRNDRFVRSRRRTGAGDCLMSLASAMDYAHRTGRALADIMRGSCLDNEFANAF-- 59
 QY 223 SWLPLDFPMTDQFGLNQESSRCYGYWKNQVIDTEGTLSHLYLHVHDGDHDKMFEC 282
 Db 60 ---PAFKPIEDIGV-----VICDQINHV----- 83
 QY 283 EGDQFICKVPMVLIVKTDNYFVPSLMLIPGD-----DELNKL- 321
 Db 84 ---SFGP-----FFPSWMNKPSTDCYVRPDEQIFQERDELNDLFGAQDCEA 128
 QY 322 -----POKATVFHILGKYLPHPTNQVGLVTRYEATLSHAD-EKGIQVR 366
 Db 129 NTVVCDACLMRCDEKAE-----RTIFRSITPRVEIDREIALQEHFHDGSHIIGVHR 182
 QY 367 VEDEDPGFQHVMDQISSCTQEKLLPEVDLVERSRHNTPKHKAIVTSLNAGYAANL 426
 Db 183 -----HGNGEDIMDHPYRWADPKVLALEQVCAIRNAKQPHSRPVRYLCLDSAAVVDHL 237
 QY 427 KSMWEYFTSGEILGIVQPSQEGYQOTEKMMN-----GALAEYLLSLTDNLV- 477
 Db 238 SGVPELFTVPKRF-----QADQAG-----PLHSAALCTDGGGSALVEMYLLGRCDVIR 287
 QY 478 ---TSAMSTF 484
 |||:|:
 Db 288 FPPTSATFTR 297

RESULT 3

Q9AMY4 PRELIMINARY; PRT; 370 AA.
 AC 09AMY4;

DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 GN NodZ.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=110SPC4;
 RX MEDLINE=21101824; PubMed=11157954;
 RA Gottfert M., Rothlisberger S., Kundig C., Beck C., Marty R.,
 RA Hennecke H.,
 RT "Potential symbiosis-specific genes uncovered by sequencing a 410-kb
 DNA region of the Bradyrhizobium japonicum chromosome."
 RL J. Bacteriol. 183:1405-1412(2001).
 DR EMBL; AF322013; AAC61006.1; -
 SQ SEQUENCE 370 AA; 42144 MW; A76834433053C58E CRC64;

Query Match 3.8%; Score 115; DB 2; Length 370;
 Best Local Similarity 20.8%; Pred. No. 0.18;
 Matches 86; Conservative 56; Mismatches 144; Indels 128; Gaps 21;

QY 120 YRKPEPYKPSVLS---KLRNYEKLHRCGPGTESYKKALKQJLDQEHIDGDECKYVW 176
 Db 4 YRCSSP-APRQAVTPGKIKRETSVLSVOPG-----REKARQWSSSSNDREYVS 55
 QY 177 ISFSGIKNRILSLASVFLYALTLDRVLVD-RGKMDDDLFCPEPLGMSWMLPLDFPMTDQ 235
 Db 56 RRRGFGDCIWSLAAAMFAKQTGTALIDWRG-----CY-----LDEPTNA 99
 QY 236 FDGLNQESSRCYGYWKNQVIDTEGTLSHLYLHVHDGDHDKMFCEGDQTFGKVPWL 295
 Db 100 FP-----VFEEVEDIG--GVAVICDD----- 120
 QY 296 IVKTDNY---FVPSLMLIPGD-----DELNKLFPKA-----TVF----- 328
 Db 121 -INRSFPFGPFPYWMNKPSPDCIYRPDEQIFEREDLDLFGSORSDANTVVCDACLM 179
 QY 329 ---HNLGRYLEFHPNTQVGLVTRYEATLSHAD-EKGIQVRFDEDPGFQHVMDQI 382
 Db 180 WRCDQEAREREIFSRPELQARLDAIYREHFEYVIGIHVR-----HONGDINGHA 234
 QY 383 SSCCTQEKLLPEVDLVERSRHNTPKH-KAVLVTSLNAGYAENLKSMTWEYFTSGEIT 441
 Db 235 PYWADTERALQIYNALDEARSLSHAKPVRAFLCTD-----SALVLEQVSVF-----DVF 286
 QY 442 GVHQPQSGYQOTEKMMN-----KALAEYLLSLTDNLV---TSAMSTF 484
 Db 287 AI--PKQFQAPQAPLHHPALGAGSGSALTETMLLAKCDTVIRPPTSATFTR 338

RESULT 4

Q65146 PRELIMINARY; PRT; 1055 AA.
 AC 065146:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 GN Helicase.
 OS African swine fever virus (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
 OX NCBI_TaxID=10497;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=21820291; PubMed=11831707;
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
 RA Rodriguez J.F., Vinuela E.;
 RT "Analysis of the complete nucleotide sequence of African swine fever

RT virus.";
RL virology 208:249-278(1995).
[12]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94233765; PubMed=8178480;
RA De la Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;
RT "Nucleotide sequence and variability of the inverted terminal
RL repetitions of African swine fever virus DNA.";
virology 201:152-156(1994).
[13]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=90219205; PubMed=2325203;
RA Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,
RT De la Vega I., Blasco R., Vinuela E.;
RL "Multigene families in African swine fever virus: Family 360.";
J. Virol. 64:2073-2081(1990).
[14]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=90219204; PubMed=2325202;
RA Almendral J.M., Almazan F., Blasco R., Vinuela E.;
RT "Multigene families in African swine fever virus: Family 110.";
RL J. Virol. 64:2064-2072(1990).
[15]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=91134988; PubMed=1994575;
RA Camacho A., Vinuela E.;
RT "Protein p22 of african swine fever virus: an early structural protein
RL that is incorporated into the membrane of infected cells.";
virology 181:251-257(1995).
[16]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX Almazan F., Murguia J.R., Rodriguez J.M., La Vega I., Vinuela E.;
RT "A set of African swine fever virus tandem repeats shares similarities
RL with SAR-like sequences.";
J. Gen. Virol. 0:0-0(0).
[17]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94187118; PubMed=8139051;
RA Rodriguez J.M., Yanez R.J., Pan R., Rodriguez J.F., Salas M.L.,
RL Vinuela E.;
RT "Multigene families in African swine fever virus: family 505.";
J. Virol. 68:2746-2751(1994).
[18]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93346971; PubMed=8339314;
RA Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;
RT African swine fever virus thymidylate kinase gene: sequence and
RL transcriptional mapping.";
J. Gen. Virol. 74:1633-1638(1993).
[19]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94065656; PubMed=8245848;
RA Alcaml A., Angulo A., Vinuela E.;
RT "Mapping and sequence of the gene encoding the African swine fever
RL virion protein of M(r) 11500.";
J. Gen. Virol. 74:2317-2324(1993).
[10]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93277388; PubMed=8503790;
RA Munoz M., Freije J.M., Salas M.L., Vinuela E., Lopez-Otin C.;
RT "Structure and expression in E. coli of the gene coding for protein
RL p10 of African swine fever virus.";
Arch. Virol. 130:93-107(1993).
[11]

RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=90357780; PubMed=2389555;
RA Blasco R., Lopez-Otin C., Munoz M., Bockamp E.-O.O., Simon-Mateo C.,
RT Vinuela E.;
RL "Sequence and evolutionary relationships of African swine fever virus
RT thymidine kinase.";
virology 178:301-304(1990).
[12]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93281390; PubMed=8506138;
RA Yanez R.J., Bournnell M., Nogal M.L., Yuste L., Vinuela E.;
RT "African swine fever virus encodes two genes which share significant
RL homology with the two largest subunits of DNA-dependent RNA
RT polymerases.";
Nucleic Acids Res. 21:2423-2427(1993).
[13]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93353606; PubMed=8102411;
RA Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;
RT "African swine fever virus encodes a CD2 homology responsible for the
RL adhesion of erythrocytes to infected cells.";
J. Virol. 67:5312-5320(1993).
[14]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94085774; PubMed=8262374;
RA Yanez R.J., Rodriguez J.M., Bournnell M.E.G., Rodriguez J.F.,
RT Vinuela E.;
RL "Two putative african swine fever virus helicases similar to yeast
RT DGR.";
Gene 134:161-174(1993).
[15]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=90223993; PubMed=2327074;
RA Lopez-Otin C., Freije J.M.P., Parra F., Mendez E., Vinuela E.;
RT "Mapping and sequence of the gene coding for protein p72, the major
RL capsid protein of African swine fever virus.";
virology 175:477-484(1990).
[16]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=94123986; PubMed=8293992;
RA Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
RT "The DNA polymerase-encoding gene of African swine fever virus:
RL sequence and transcriptional mapping.";
Gene 136:103-110(1993).
[17]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93327788; PubMed=8335009;
RA Simon-Mateo C., Andres G., Vinuela E.;
RT "Polyprotein processing in African swine fever virus: a novel gene
RL expression strategy for a DNA virus.";
EMBO J. 12:2977-2987(1993).
[18]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=93233210; PubMed=8474154;
RA Prados F.J., Vinuela E., Alcaml A.;
RT "Sequence and characterization of the major early phosphoprotein p32
RL of African swine fever virus.";
J. Virol. 67:2475-2485(1993).
[19]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE=92260660; PubMed=1583732;
RA Alcaml A., Angulo A., Lopez-Otin C., Munoz M., Freije J.M.P.,
RL Carrascosa A.L., Vinuela E.;

RT "amino acid sequence and structural properties of protein p12, an
RT African swine fever virus attachment protein.",
RL J. Virol. 66:3860-3868(1992).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93174976; PubMed=8438592;
RA Yanez R.J., Vinuela E.;
RT "African swine fever virus encodes a DNA ligase.",
RL Virology 193:531-536(1993).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93174941; PubMed=8382399;
RA Pena L., Yanez R.J., Revilla Y., Vinuela E., Salas M.L.;
RT "African swine fever virus guanylyltransferase.",
RL Virology 193:319-328(1993).
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=95159428; PubMed=7856088;
RA Simon-Mateo C., Freije J.M., Lopez-Otin C., Vinuela E.;
RT "Mapping and sequence of the gene encoding protein p17, a major
RT African swine fever virus structural protein.",
RL Virology 206:1140-1144(1995).
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=92263807; PubMed=1316688;
RA Garcia-Beato R., Freije J.M.P., Lopez-Otin C., Blasco R., Vinuela E.,
RA Salas M.L.;
RT "A gene homologous to topoisomerase II in african swine fever virus.",
RL Virology 188:938-947(1992).
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=94091056; PubMed=8266720;
Query Match 3.8%; Score 114; DB 12; Length 1055;
Best local Similarity 19.2%; Pred. No. 1.1; 214; Indels 134; Gaps 25;
Matches 104; Conservative 90; Mismatches 134; Indels 134; Gaps 25;
QY 15 TTGGSKSVNSELLO-MKYLSSGTMKLTFTTCLIVFSVLAFA-----SMIFQH 65
DB 257 TTPSSCDTIQOSELEVKF--AGTLCKNHF--LRVHNLVLFKRTSPSYCEICKRMH 311
QY 66 PPSNRMGAELARVLDAGFVPVNTINSKLLGGLA--SGFDE--DSCLSR---YQ 116
DB 312 DKNTLTL-----RVGNKYYQHCGRHDKHSLMGSLSGTTFEVEYVVDQVMTKSIEVHE 366
QY 117 SVHYRKPSPKPSYSILSKLRNFE-----KLHRCGPETESYKAKLQLDGHDG 169
DB 367 SLTFEELPTQKHTIDESSKRETERVFTLVYKAKMKIGKTVQLRNLQKYGNNSISKQ 426
QY 170 ECKYVW--TSFSGLRNLTSLASVFLYALLT-----DRVLLVDRKDDDFCEPFLG 221
DB 427 TIRFVTFEQIFSKNIQSR--LPNFTLYSEVTDLDLSYEIVY-----QVESL-----FRL 474
QY 222 MSMLPLDPFMTQDFGL-NOESSRCYGM-----VKNOYIDTEGTLSHLYH 268
DB 475 TSTAEPDILLIDEVESIFNQFNSGLHKYFAPSFALFMMLLETANVYICIDAMGNRTYN 534
QY 269 LVHVDY-GD-----HDKMFECGDOTFICKVWMLVKINDNFVPSLMLIPG- 312
DB 535 ILDFRQDVPITFHMNOYKRAQHDITYFTSSRETWLNLLKDLLEOKKIVIPNSLMEAR 594
QY 313 -FDDELKLFEPKATVFN-----HLGRYLFHPTNOVGLV-----TRYEAVLSHADEK 360
DB 595 ILQSPQKPKREKKIGFYSKSTAHRESHFNVSYWGLVDLLITPTISAGVSIEDKR 654
QY 361 IGIVQVFEDEDPGFQHVMDQISSCQKEKLPEDVTLVRSRHVNTPKHKAVALVTSNA 420
DB 655 -----FDVLYGTFNNMSCDVEYTCQ-----MLGRVRELKSKCYKICLQCKQNY 697

QY 421 GVAENLK-----SMYKEYPTSGELIGVHOPSQEQOTEKKMN 460
DB 698 -YPTTIEDIMFLQKRDITLFTIINNHOLSFTYSKETGRPVYKTPPYHMLTMRIOHL 756
QY 461 GK 462
DB 757 SK 758
RESULT 5
Q9ZM93 PRELIMINARY; PRT; 2149 AA.
ID Q9ZM93
AC Q9ZM93;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-DIC-2001 (TReMBLrel. 19, Last annotation update)
DE F5A8.5 protein.
GN F5A8.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim G., Li Y.,
RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,
RA Yu G., Ecker J., Theologis A., Davys R.W.;
RL Submitted (JAN-1999) to the EMBL/Genbank/DBD databases.
SR EMBL: AC004146; AAD10656.1; -
SQ SEQUENCE 2149 AA; 237744 MW; BDE94CF7FC307B9B CRC64;

Query Match 3.7%; Score 112.5; DB 10; Length 2149;
Best local Similarity 23.6%; Pred. No. 4.2;
Matches 123; Conservative 65; Mismatches 186; Indels 147; Gaps 28;
QY 52 FSVLVAFSMTIHFHPSDSNRIMGFAEARVLDAGVFP-NVTNINSKLT-----L 98
DB 1351 FAVLVLFQ--GMHPASSMQVLPFSSAKIKSGDSSINSSFGGLKYRLALPVFQSLSA 1407
QY 99 GGLLASGF-DEDSCLRYQSYNHRKPSPPKSSYLSKLRNEKHLKRCGPETESKRL 157
DB 1408 GRFFSSGFSLIDCQELDLYLST---SFHDSMDTLAVSVQOL-----VGTKIYYSFV 1459
QY 158 KQLDQEHIDGDECKYVWVISEFGLGNR-----ILSLASVFL-----YALTDRVLL 204
DB 1460 -----GBLISPCFCYPAIGISPDIMFMWITPLFDFELRLVMPYALEKKRYL- 1507
QY 205 VDRGKMDLFCBPFLGMSWLLPLDFPMTQDFGLNQESSRCYGMW-----NOYIDTE 259
DB 1508 -----MKFTF--FOQISQNCPRKDFLESEFAVSTIE--LCIGYLFKILHHRNEISPD 1556
QY 260 G-----TLSHLYL-----HLVHDYGDHDKMFCEBGQTFIKGKPMILVTDNTPVS 306
DB 1557 GIWDMNLSPFLFISIKTLVKRFELKRLNLSAPLAFLLSGKVC-IKQVP-----TDALPKA 1610
QY 307 LMLIPGEDD--ELNKLFPQK-----ATVFHHLGYLFLHPTNOVGLVTRYEAVL 354
DB 1611 LEIVKSTNDLLELTPRASSQKPYTOSTNFAADSGPHL-RAIFGACLIHMGDLTR----- 1663
QY 355 SHADEKIGIVQVFEDEDPGFQHVMDQISSCQK-----EKLPEVDTLVRSRHVNTPKH 410
DB 1664 ---DCINGIQL-VDSKRSGLRKLDLKLVECLEQVLSLAKLAYEFDCPVDER--NTNSI 1716
QY 411 KAVLYTS-----LNAGYAEMLKSMYWEYTS-----TEGIIQVHOP 446
DB 1717 CIYMLKSCQISIAAVVKSQNVQVATVQLVLSLQVRVNNPEKSFVILFVGLIGDIXP 1776
QY 447 SOEGYQOTEKKMNGKALAEWYLL--SLTDNLVTSANSTF 484

```

Db 1777 -----VFESVVIAGECLRFIMLQTHSITDELQKGFMSLF 1812
RESULT 6
ID 0997F0 PRELIMINARY: PRT: 2244 AA.
AC 0997F0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Polymerase (RNA polymerase).
GN L.
OS Nipah virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae.
OX NCBI_TaxID=121791.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20287714; PubMed=10827955;
RA Chua K.B., Bellini W.J., Rota P.A., Harcourt B.H., Tamlin A., Lam S.K.,
RA Ksiazek T.G., Rollin P.E., Zaki S.R., Shieh W., Goldsmith C.S.,
RA Gubler D.J., Roehrig J.T., Eaton B., Gould A.R., Olson J., Field H.,
RA Daniels P., Ling A.E., Peters C.J., Anderson L.J., Maly B.W.;
RT "Nipah virus: A recently emergent deadly paramyxovirus.";
RL Science 288:1432-1435(2000).
[2]
RP SEQUENCE FROM N.A.
RA Harcourt B.H., Tamlin A., Halpin K., Rollin P.E., Ksiazek T.G.,
RA Bellini W.J., Rota P.A.;
RT "Molecular Characterization of the Polymerase Gene and Genomic Terminal
RT of Nipah Virus.";
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RX STRAIN=UMMC2;
RC MEDLINE=21405817; PubMed=11514724;
RA Chan Y.P., Chua K.B., Koh C.L., Lim M.E., Lam S.K.;
RT "Complete nucleotide sequences of Nipah virus isolates from
RT Malaysia.";
RL J. Gen. Virol. 82:2151-2155(2001).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=UMMC2;
RA Chua K.B., Chan Y.P., Koh C.L., Lim M.E., Lam S.K.;
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF212302; AAK28089.1; -
DR EMBL: AY029768; AAK50555.1; -
DR InterPro: IPR001016; Viral_RNA_pol_L.
DR Pfam: PF00946; Paramyx_RNA_pol.1.
SQ SEQUENCE 2244 AA; 257230 MW; 1DE123E651A738C9 CRC64;

Query Match 3.7%; Score 112.5; DB 12; Length 2244;
Best Local Similarity 18.8%; Pred. No. 4.5;
Matches 104; Conservative 79; Mismatches 192; Indels 177; Gaps 25;

OY 1 MDONSYRRSSPIRTTGGSKSVNFSELLQMKYLSGTMKL--TRFTTCLIVFSVLAVF 58
Db 1804 VENHKYRRIG---LNSSCYKALNLSPLIQ-RILPSGAORLFGESGSMMLLYOSTLQ 1859
OY 59 SMIFHQHPDSNRIMGFARVLDAGVFPVNTNINSK-----LLGLLASGFDESDCS 113
Db 1860 SISFYNSGIDGYIPQORELK-----LFPSEVSIADPSLTGKLGVLVPLFNG----- 1909
OY 114 RQSVHYRRKPSYKPSYLSIKRNEYEKLHKRCGPTESEYKALKQDDEHIDGDECKY 173
Db 1910 -----RPETTWIGNDSTEYIINR---TAGRSIGLVSDME----- 1942
OY 174 VVWISFSGLGNR-----LSLAVFLYALTDRLVLDVRGDMDDLCEPFLGMSWL 226
Db 1943 -----SGIDKNVEELIVESHLSIAINVMMDGL-----VSKIATYTPGFPI SRL- 1988
OY 227 PLDFPMTDFDGL-----NQESSRCYGYVKNQV---IDTECTLHLVLR-LVHDY 273

```

```

Db 1989 ---FNNRYSYFGLVLCFPPVYSNPDSDEVYLLCLOKTVKTIYPPQKVLBESHNLHDEVNDQ 2045
OY 274 GDHDKMFPEGQDTFICGKVPWLIVKTDNFEVPSMLILPEPDEDLNLFQKAVFHHGLR 333
Db 2046 GITSVLEFKINSQS-----KQFHDLKR-----VYQIDQ 2074
OY 334 YLFHPT-----NQVW---GLVTRYEAYLSHADEKIGIGRV-----FDE 370
Db 2075 PFEVPPKITSDEYVLLQAGIKLNGPRLKSEISYDIDGSDINTLRDTIITMLNFAAMYFDD 2134
OY 371 DPGPFQHV-----MDQISSCTQKE---KLPEVDTLVERSHVNTPRKAVLVY 416
Db 2135 NRSPSHHLEPPYVLEKTRIKTINVCVKTKYIVSLIKPDKTSSELYHKNNIRKVLTI 2194
OY 417 SLNAGYAENLKSMTWEPYPTSGEIIIGVHPSQCYOOTEKKMHNGALAEMLYLSLTDNL 476
Db 2195 DFRS-----KLMTKTLP-----KQMOERREK--NG--FKEWIVLDSNRE 2230
OY 477 VTSAMSTEGYVA 488
Db 2231 VKIWKIITIGIIS 2242

RESULT 7
ID 091AE5 PRELIMINARY: PRT: 2244 AA.
AC 091AE5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE RNA polymerase.
GN L.
OS Nipah virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae.
OX NCBI_TaxID=121791.
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=UMMC1;
RC MEDLINE=21405817; PubMed=11514724;
RA Chan Y.P., Chua K.B., Koh C.L., Lim M.E., Lam S.K.;
RT "Complete nucleotide sequences of Nipah virus isolates from
RT Malaysia.";
RL J. Gen. Virol. 82:2151-2155(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=UMMC1;
RA Chua K.B., Chan Y.P., Koh C.L., Lim M.E., Lam S.K.;
RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY029767; AAK50546.1; -
DR InterPro: IPR001016; Viral_RNA_pol_L.
DR Pfam: PF00946; Paramyx_RNA_pol.1.
SQ SEQUENCE 2244 AA; 257290 MW; 993DC4782846E268 CRC64;

Query Match 3.7%; Score 112.5; DB 12; Length 2244;
Best Local Similarity 18.8%; Pred. No. 4.5;
Matches 104; Conservative 79; Mismatches 192; Indels 177; Gaps 25;

OY 1 MDONSYRRSSPIRTTGGSKSVNFSELLQMKYLSGTMKL--TRFTTCLIVFSVLAVF 58
Db 1804 VENHKYRRIG---LNSSCYKALNLSPLIQ-RILPSGAORLFGESGSMMLLYOSTLQ 1859
OY 59 SMIFHQHPDSNRIMGFARVLDAGVFPVNTNINSK-----LLGLLASGFDESDCS 113
Db 1860 SISFYNSGIDGYIPQORELK-----LFPSEVSIADPSLTGKLGVLVPLFNG----- 1909
OY 114 RQSVHYRRKPSYKPSYLSIKRNEYEKLHKRCGPTESEYKALKQDDEHIDGDECKY 173
Db 1910 -----RPETTWIGNDSTEYIINR---TAGRSIGLVSDME----- 1942
OY 174 VVWISFSGLGNR-----LSLAVFLYALTDRLVLDVRGDMDDLCEPFLGMSWL 226
Db 1943 -----SGIDKNVEELIVESHLSIAINVMMDGL-----VSKIATYTPGFPI SRL- 1988

```

```

0Y 227 PLDFEFTPOEFGI-----NOESSRCGXWKNQV---IDFGTSLHLX- LVNDY 273
Db 1989 ---FMNYSYGLVJWCFEPYVSNPDSIEVLLCLOQTVTYTPPOKVEHSHLHDEVNDQ 2045I
0Y 274 GDHDKMFCBGSDQTFICKVFWLIVKTDNVPFVPSLWILPGFDELNKLPFOKATVFNHNGR 333
Db 2046 GITSVIEFKIKNSO-----KOFHNDIKR-----YYQIDQ 2074
0Y 334 YLFHPT-----NQW---GLVTRYEAYLSHADEKIGIOVR-----FDE 370
Db 2075 PEFVPTKTSDEQYLLADQKINGPELKSSEISDIQSGJNLRPTIILMEANNYDD 2134
0Y 371 DPGFQHV-----MOISSCQKE---KLPEVDTLVSRHVNTPKRAVLVT 416
Db 2135 NRSPSHLEPEYLERIRIKTINNCVTKRYVSLIKFEDTKSSSELYHKNNIRKRVIL 2194
0Y 417 SLNGLVAENLKSMTWEYPTSTGEIIGVHOPSGEYQOOTEKKHNGKALAEWYLSITDNL 476
Db 2195 DFRS-----KLMKTKLP-----KGMQERREK--NG--FREWYVDLSNRE 2230
0Y 477 VTSAMSTFGYVA 488
Db 2231 VKIWMKIIIGYIS 2242

```

RESULT 8		
ID	ID	PRT; 322 AA.
085712	PRELIMINARY;	
AC 085712;		
DT 01-NOV-1998	(TREMBLrel. 08.	Created)
DT 01-NOV-1998	(TREMBLrel. 08.	Last sequence update)
DT 01-DEC-2001	(TREMBLrel. 19.	Last annotation update)
DE	Modulation protein Nod2.	

OS Rhizobium fredii (Sinorhizobium fredii).
OG Plasmid pSFHH103d
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_taxid=380;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HH103;
RX MEDLINE=99164952; PubMed=10065558;
RA Lamberth Y., Bellagin R.A., Cudo T., Espuny R., Gill A., Krishnan H.B.,
RA Megias M., Olivero F.J., Pueppke S.G., Ruiz-Sainz J.E., Spark H.P.,
RA Tejero-Mateo P., Thomas-Oates J., Vinardell J.M.;
RT 'Mutation in GDP-fucose synthase genes of Sinorhizobium fredii alters
RT nod factors and significantly decreases competitiveness to nodulate
RT soybeans.';
RT MOL. Plant Microbe Interact. 12:207-217(1999).
DR EMBL: AF072888; AAC27749.1; -.
KW Plasmid.
SQ SEQUENCE 322 AA; 36535 MW; 0023D958F17781CB CRC64;

```

Query Match      3.7%; Score 110.5; DB 2; Length 322;
Best Local Similarity 20.2%; Pred. No.0.377;
Matches 74; Conservative 51; Mismatches 109; Indels 133; Gaps 17.

QY   172 KYVWISSGIGNRILSLASVFYALLTDRVLVD-RCKDMDFCEPFLGSMLPLDF 230
      ||| : : : : | : | : | : | : | : | : | : | : | : | : | : |
Db    4 RYVLSRRRTGGDDCWSLAAMRYAORTARLTAVDWRGS-----CY-----LDQ 47

QY   231 PMTDQFDLINOESSRCXYMKVNOVIDTEGTLSHYLVHVDYGDHDKMFCEGDQTFIG 290
      ||| : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    48 PFTNAFP-----VFEEPIKD-----LA 64

QY   291 KVPMTIVKTENVY-----FVPSLMILPGBD-----DELNKLFEOKARY- 327
      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    65 GVPEI---CDNQINIEFSPPGFPMWMKKPALIECYRRDAOVFRERDELDLPQAODDVE 121

QY   328 -----FHHLGRLYF---HPTNOVWGILVTIRYEALYSHADEKIGIOVRVED 370
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Db 122 ANTVCCGACGLMMRDEBEAERQIFCSVPRALIQARIDAIYQEH-YQSAIGHVR---- 176
QY 371 DPGPQHVMDIOISSCTQXKILLPEVDLTVLRSRHNVPKHKRAVLVLSLNGAENLKSUY 430
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 -HGNGEDVMDHAPYAWADPDLLVHQVCAINAKALLPRKPKPVRYLLCTDSARVLDDQYSSRF 235
QY 431 WEYPL-----STGELIGVHDPDSQSGYQOTEKKMHNG-KALAEMLLSLTPNLV---- 477
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 PDLLLIPIKSRFADSGPL---HSP-----DLGEGGTSALVEVMYLLGICPTVIREPP 284
QY 478 TSAMSTF 484
      : : : : :
Db 285 TSAFTRY 291

```

RESULT	9		
Q21310			
ID	Q21310	PRELIMINARY:	PRT; 529 AA.
AC	Q21310;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	k08C7.5 protein.		
GN	k08C7.5.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;		
OC	Rhabdilitidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
UN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Berks M.;		
RL	Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	none:		
RT	"Genome sequence of the nematode C.elegans: A platform for		
RT	Investigating biology.";		
RL	Science 282:2012-2018(1998) .		
DR	EMBL: 270286; CAA94290.1; "		
DR	InterPro: IPR000960; Flav_conf_mmoxn.		
DR	Pfam: PF00743; FMO-like; 1		
SO	SEQUENCE 529 AA; 60140 MW; 05EDD3446D3FAAC1 CRC64;		

	Query Match	3.7%;	Score 110;	DB 5;	Length 529.	
	Best Local Similarity	20.3%;	Pred. No. 0.85;			
	Matches 102;	Conservative	78;	Mismatches 142;	Indels 180;	Gaps 30.
QY	73 MGRFAERVLADACV--PPNVTININDSKLLGGLASGFEDSDC-----LSRYOSVHYR	121				
		:	:	:	:	:
	1 MGKKRAVAIVAGASGLPSIRH-----GLL-YGFVDV-TCFEASDDIGGIAMRYKSHETN	50				
Db						
QY	122 KPSPYK-----PSSYLISKLRNVKRLHKRCGGTSTSYKALKOJDOEH	164				
	: : :	:	:	:	:	:
Db	51 ESSVAKTYIINTSKEMTAVSDFTPQENIANPMHNMEMI-----NYPKS-----Y	94				
QY	165 IDDGSECKRYWVISFSGIGNRIILSLAVELALLTDRLVLWDGKDMDJFCSPFIWSM	224				
	: :	:	:	:	:	:
Db	95 AEHHGLMKHI-----KLRRHLNIERSKNYN-----DGTW	125				
QY	225 IL-----PLDFPMTDQFDGINQESSRCYGVV-----KNQVIDTECTSLHYLIHV	270				
	: : :	:	:	:	:	:
Db	126 KATVIOPEKTELTLELFDDV---LVCSGHNAIPHWPRPFQONDF---KQRIYH-----S	173				
QY	271 HDYGDH---DKMPCEG-----DQTFIGKVFWMLIVLKIDNFVFESLWL--IPGFD	314				
	:	:	:	:	:	:
Db	174 HDVKDHKGVEDVVVVVVGISNGSIDVAEQSRIKQVUYLVTRRGSTMILIKRELERGIP-FD	232				
QY	315 DELN-----KLFQO---KATVEFHNLGRVLEHPTNOVGC--VTRYEAUYSHADE--	359				
	:	:	:	:	:	:
Db	233 IIMNTFFSLYKLPFOQALMNSLIVERLNQRIDH---DLVGLKAHHNVFSAHPSLNDDELPN	289				
QY	360 -----KIGIOVRVD-----EDPGFPQHVMDOISS-----CTQKEKLLPEVDTL	398				

```
Db 290 RIANGTVRIKPIKPKFDYAIHFEDGTVPHVDEYVMSTGSEFENLIEHGKLVSENE 349
Oy 399 VERSHH---VMTPKRAVLVTSINAGY-----AENLSKMYVEYFTSGEILGVQPSQE 449
Db 350 VDLFYMFPVATSDNSNLCIGLIPDGSIMPVSQOARVEFPANNVSGNNLI-----PKRS 405
Oy 450 GYOQTEKRMHNGKALAEWYLLS 471
Db 406 --QMSEDVLNKKREMAAQOFVKS 425

RESULT 10
O25252 PRELIMINARY; PRT: 658 AA.
ID 025252
AC 025252:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein HP0513.
GN HP0513.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomo J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khaliq M.G., Glodek A.,
RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL: AE000565; AAD07579.1; -
DR TIGR; HP0513; -
DR InterPro; IPR004919; DUF262.
DR Pfam; PF03235; DUF262; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 658 AA; 78181 MW; 72CF7028B68912D9 CRC64;

Query Match 3.7%; Score 110; DB 16; Length 658;
Best Local Similarity 17.8%; Pred. No. 1.2;
Matches 102; Conservative 79; Mismatches 191; Indels 202; Gaps 23;

Oy 37 GTMLTRFTTCLIVFSVLAFSMIFHOHPSDSNRIMGFAARVLADGVFPNVTININDK 96
Db 28 GQQRILT-TITLLLA-----LRNHLSEVEILEKFSRKELESYLIN--SKDDDK 74
Oy 97 LLGILLASGFEDSCLSRQSVHYARKRPSYKSSYLLSKLNRYEKLHRCBPGTESYKKA 156
Db 75 KF-RLLISESKDQLLSLIDK-NKRKPS--EPSYKIVNFELEKEMISENDKLETFIKG 130
Oy 157 LKOLDQEHIDGDECKYVWIS-----FSGLNRLISLASVPLVALLDRLVLLV 205
Db 131 LKRL-----MIVISLIDKGDQDPLIFESMNSKOIELTQ--DLIRNYIME 175
Oy 206 DRGDMODLCEPFLGSMWLPRLDFPMIDQF----- 236
Db 176 TEVEQOEDEFYQVYRAMEREQNETLENRVRVYLTIKIGIKINERKRYEAFKDYROK 235
Oy 237 ---DGLNOESSRCGYWV-----KNQVIDTEGTLSH-----LYLHVHDYGDH 276
Db 236 GIEIEDLLKLDQKICGVCQIAFKKEDDKLANKLSFLVLEMDVITYPLLELSDYKD- 294
Oy 277 DKMFCEGDQTFIGKVPWLIVKTDNYFVPSLMLPGF-----DDELINKLPF----- 322
```

```
Db 295 -----GLSKQD--FIIPIILISYICRAVCGLGINSLNKVPSPFTK 335
Oy 323 --QKATVPHHLCRYLF-----HPTNOYGLV-----TRYEAYVSHADEKIG 362
Db 336 HIOKDEYFKSLKAHFVCLTEKORFPNDEFFKFLITIDYFKFKKKKFLERLENDQKEP 395
Oy 363 IQVRFEDEDPGFQVHVMQIISCTQKEKLLPEVDLVERSRIV--NTRKRAVLVTSLSLA 420
Db 396 V-----DTQKCNIEHIMQTLTPEMQRDGENFQAIEKYLHTIGN-----LTLTGYS 444
Oy 421 GYAEN-----LKS-----YMEYPTST 437
Db 445 KYSNNSFQEKRMKMGKFGKSSLSKLQSLKDLSEFGKEIEKRAADLADMAKIWTPILE 504
Oy 438 GEIVGHQPSQEGYQOTEKRMHNGKALAEWYLLS 471
Db 505 AETLEEKPKKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 538

RESULT 11
O07183 PRELIMINARY; PRT: 1102 AA.
ID 007183
AC 007183:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE Putative origin of replication binding protein lwm6dL.
OC African swine fever virus (ASFV).
OS Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfarvirus.
OX NCBI_TaxID=10497;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=93275762; PubMed=8389041;
RA Susman M.D., Lu Z., Kutish G.F., Alfonso C.A., Rock D.L.;
RT "The identification of an African swine fever gene with conserved
RT helicase motifs and a striking homology to herpes virus origin binding
RT protein, UL9."
RL Nucleic Acids Res. 21:2254-2254(1993).
CC - FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
DR EMBL; L12174; AAA42719.1; -
KW DNA replication; DNA-binding; Helicase.
SQ SEQUENCE 1102 AA; 129255 MW; 4B06FA67304C60 CRC64;

Query Match 3.7%; Score 110; DB 12; Length 1102;
Best Local Similarity 19.0%; Pred. No. 2.6;
Matches 103; Conservative 94; Mismatches 210; Indels 136; Gaps 26;

Oy 15 TTTGSKGVNFSSELD--MKYLLSGTGMKLTFTTCLIVFSVLAF-----SMIFHOH 65
Db 251 TTSSSCDSIOOSELEEVLF--AGTLCKNHCF--LRVYKNIIVLEKRTSPSYCELCIKRMH 305
Oy 66 PSDSNRIMGFAEARVLADGVFPNVTININDKLLGLSLA--SGFDE--DSCLSR-----YQ 116
Db 306 DKDNTFLI-----RVTGKKVYQHCNDKHSLLMSLSGTTNFTVYEQVMSKIEVNE 360
Oy 117 SVHYKRPSPYKSSYLLSKLNRYE-----KLHKRCBPGTESYKKAOLKOLDQEHIDG 169
Db 361 SILFEELPDQTKHYDESSMREYERVPLLVKAKQIKGTIQ-----LRNYLQKYGDS 415
Oy 170 ECKY--VVMISFGSGNLTLS--LASVFLYALTT-----DRVLLVDGKMDMDLFCPEF 219
Db 416 ISKQOTIRFVTFROTFSSNIQTRLPNFTLYSEVTDGDDLSYERVII--OVESL---F 466
Oy 220 LGMWLPRLDFPMIDQFQGL--NOESSRCGYWV-----VKNOVIDTEGTLSHLY 266
Db 467 RLTSABRPVLDLIDVESIFNQFNSGLHKYFAPSFALPMMLERANVITCLDANLGKRT 526
Oy 267 LHLVHDY-GD-----HDKMFCEGDQTFIGKVPWLIVKTDNYFVPSLMLIP 311
Db 527 YNIQRFRGDVPIFFHNMQYQAKQNDMYFTSSREIMWLNILKDLLEDKKIVIPNSLME 586
```


RESULT 14
094648 PRELIMINARY: PRT: 3844 AA.
AC 094648;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE AARPL protein (Fragment).
CN AARPL.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN NCBI_TaxID=36329;
RX MEDLINE=937378065; PubMed=9234746;
RA Barale J.C., Candelle D., Altal-Bonnefoy G., Dehoux P., Bonnefoy S.,
RA Barale J.C., Pereira da Silva L., Langsley G.;
RT "Plasmodium falciparum AARPL, a giant protein containing repeated
RT motifs rich in asparagine and aspartate residues, is associated with
RT the infected erythrocyte membrane."
RL Infect. Immun. 65:3003-3010(1997).
DR EMBL: Y08926; CAAT0130.1; -
DR InterPro: IPR001841; znf_ring.
DR SMART: SM00184; RING; 1.
FT NON_TER 1
FT SEQUENCE 3844 AA; 456075 MW; 23D0233261C30DBB CRC64;
SQ

Query Match 3.5%; Score 106; DB 5; Length 3844;
Best Local Similarity 18.1%; Pred. No. 37;
Matches 91; Conservative 60; Mismatches 131; Indels 220; Gaps 24;

QY 23 VNESELLQMKYLSSG-TWKLTRFTTCLIVSVLVASMFHQPDSNRIMGFAEARVL 81
DB 371 INOKRIILOERENGNMKNIT-----LFHFKEKF-----TSN-----DDIYL 407
QY 82 DAGVPPNTNINSKLLGGLASGFDESDCLSRQSVHYRKPSPYK-----PSSYLK 135
DB 408 ESNYFKMTMMNNSSKLL-----RECNNKYR-IKYRETIQEKNIKIKCKNIYERK 456
QY 136 LRN--YEKLHRCGPGTESYKALKQLDQEHIDGDECKYVWISFGLNRIILASVF 193
DB 457 LKNSKYE-----FNEFLD-----IKY----- 472
QY 194 LYALLTRVLLVDRGKMDLFCPEFLGMSWLLPLDFPMTDQFDGLNDESSRC----- 246
DB 473 -----DKEKNITLHP-----LYIDLVMGDIFFELNLFDLKLNILFI 511
QY 247 -----YGYMKANQVYIDTBGTLSHLYLVHDYGDHDKMFCFEGQOTPIGKVPW 294
DB 512 NIQLTFEYIEIKYQWANGP-----OMIFQVGEYDLSLFFQND--LIG-IQF 556
QY 295 LIVATDNVYFVSLMLPFGDELNKLFPQKATVFNHLGRYLFHPTNOYWGVLTRYEA-- 352
DB 557 ALIMMN-----VLPLEYENRKYVLDLSHVYLNTH-----QIMNKKVQYDEPLK 597
QY 353 -----YLSHADEKIGIOVRVDEDPGPFQHVMDQISSCTQEKLLPEVDY 397
DB 598 KLPMWSKNEYNDQVMDYIHKRDEDEICINKEYDEDTNKKY-----IDT 638
QY 398 LVESRHHVNTPKKRAVLVTSNAGYAENLKSMMY--EYPTSGEITIGVHQPDSOG--Y 451
DB 639 SPNFSYNHNIQNHVYV-----EEHSEDKPKPYMNKIKYIKKNDFFEEHKKMYESMLIY 692
QY 452 QQTEK-----KMHNGK 462
DB 693 NRSKKSNDNNTLPLVNNKMGNSK 714

RESULT 15

001660 PRELIMINARY: PRT: 381 AA.
AC 001660;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 44.4 kDa protein.
GN T28F2.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitida; Peloderinae; Caenorhabditis.
RN NCBI_TaxID=6239;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Madsen C., Fronick B.;
RT "The sequence of C. elegans cosmid T28F2."
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF000198; AA853053.1; -
DR InterPro: IPR002516; GT_11.
DR Pfam: Pf01531; Glyco_transf_11; 1.
KW Hypothetical protein.
SQ SEQUENCE 381 AA; 44370 MW; 502A9742034053C1 CRC64;

Query Match 3.5%; Score 105; DB 5; Length 381;
Best Local Similarity 18.3%; Pred. No. 1.4;
Matches 67; Conservative 49; Mismatches 97; Indels 154; Gaps 18;

QY 296 IVKTDNYPVPSL-----WLIPGF-----DDELNKLFPQKATVF 328
DB 11 ILKNNKEFYIIVLLVIFOMWPGYDYNOISTLYYCDKNPSDQRIYLFPRVTVTF 70
QY 329 H-HLGRYLFHPTNOYGVV-----TRYV--EAYLSH-----ADEKIGIQ 364
DB 71 DGLGNQLF-----EYFSLGLALKLNRPATFNSEDMILSKNLLRQGVQVAERVISIP 126
QY 365 VRV-----FDEDPGPFQHVMDQISSCTQK-----KLLPEVDLYERSRHVNT 407
DB 127 IEIPSSRFYSPACHPQFSPSLSCESQKYLVIDGHYFQSHKYPSSISITSIR--KWLNP 184
QY 408 PKHKA-----VLVTSNAGYAEV-----LKSWMYEP-- 434
DB 185 PQBEKLLKMKIRKDERFKICVHIRGDFLTDSQAHGTSNFTIRAVDHLTYOHPGLV 244
QY 435 ---TSTG---EIIGVHQPDSOGYQOTEKMHNGKALAEMLLSTLDN--LVTSAMSTFG 485
DB 245 YLFSDNPGWREKITAHL-----DYQSDVKVMTSEAIKIDLYFAQIHCDAYLITAPSTFG 300
QY 486 YVAQGLGLKPMILYRPNRTTPRPSGCRASMECFHSPFPPYDCKATGTGIDTGLVHVY 545
DB 301 W-----WLGYSKNSQSSQTPNGR-----GFLPVHV 326
QY 546 RHCEDIS 552
DB 327 EIGIDVT 333

Search completed: February 27, 2003, 15:31:50
Job time : 46 secs

```

: APPLICANT: GOODRICH, LISA V
: APPLICANT: JOHNSON, RONALD L
: TITLE OF INVENTION: Patched Genes and their Use
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Foley, Hoag & Eliot
: STREET: One Post Office Square
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII(text)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/954,701A
: FILING DATE: 20-OCT-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Vincent, Matthew P.
: REGISTRATION NUMBER: 36709
: REFERENCE/DOCKET NUMBER: SUV-003.08
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-832-1000
: TELEFAX: 617-832-7000
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1285 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-954-701A-6

Query Match 3.0%; Score 89; DB 1; Length 1285;
Best Local Similarity 20.3%; Pred. NO. 17;
Matches 63; Conservative 39; Mismatches 83; Indels 126; Gaps 16;

OY 218 PFLGSMW-----LLPLDPRMIDQFDGLN-----QESSRCYGYM 250
    ||| ||| : : : ||| : : : |||
Db 672 PFLMSWVKFLTWGFLAALISSLYASTRLDGLDIIIDLPKDSNKHKFLDAQTLFGRY 731
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 251 VKNQVITDGLTSLH-LYLHLVHDYGDHDKMEFCEGDQFFIGKVPWLIVKTDVYFVPSLWL 309
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 732 SMYAAV--TQGNREYTYTQQLLRDY--HDSF-----RVPH-VIKNDNGGLPDEFWL 775
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 310 IPGEDDELNKLFPQKATVYFHHLGRLYLFHPPTNMGWLVTRYEAYLSHADEKIGIQRVFD 369
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 776 L-----LF-----SEWLGNIQKIFD 790
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 370 EDPGPFQIVMDQISSCTQKEKLLPEVDI-----LVERSRHVNTPKHKAIVYS--LNA 420
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 E-----EYRQGRLTKECFHPASSDAILIAYLLIQTGHVNDPVDKELVTLRNLVNS 841
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 421 GYAENILKSMY-----WEYPTSTGELIGVHQPSQEGYQTEKKMHNGKAL-----A 465
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 842 DGIINQRAFYNNILSMATINDVFAVYASQSKL--YEPRIQYFHQRPNEYDLKIKPSLPVYA 899
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 466 EM--YLLSLTD 474
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 900 QMPFYHLGLTD 910
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-10-041-030-4
: Sequence 4, Application US/10041030
: Patent No. US20020150934A1
: GENERAL INFORMATION:
: APPLICANT: Powers, Scott
: APPLICANT: Mu, David
: APPLICANT: Xiang, Phil

```

```

: APPLICANT: Peng, Yue
: APPLICANT: Toutarik, Inc.
: TITLE OF INVENTION: Diagnosis and Treatment of Cancer Using Mammalian
: TITLE OF INVENTION: Bellino Polypeptides and Polynucleotides
: FILE REFERENCE: 018781-00681005
: CURRENT APPLICATION NUMBER: US/10/041,030
: CURRENT FILING DATE: 2001-12-28
: PRIOR APPLICATION NUMBER: US 60/259,502
: PRIOR FILING DATE: 2001-01-02
: NUMBER OF SEQ ID NOS: 42
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO: 4
: LENGTH: 420
: TYPE: prt
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: human pellino 2
: US-10-041-030-4

```

Query Match	2.9%;	Score 88.5;	DB 12;	Length 420;
Best Local Similarity	19.6%;	Pred. No. 3.4;		
Matches	81;	Conservative	52;	Mismatches 120;
			Indels	161;
			Gaps	24;

```

0Y 87 PNVNINSDKLL-----GGLASGDESDCSLRXOYVHKRSPKPPSS-YLISLRYEK 144
Db 12 PKKEPVYKIGELVVLGYNNALPNG-DGRKRSRFALYKRPKANGVKPSVTHIS----- 63
0Y 142 LHKRCGPSTESYKRAKLOLDEHIDGDECEQYVWVIFSGJGNRIILSLAVFALLTDR 201
Db 64 -----TPQASKAI-----SCR-----GQHSIS-----YTLSRNQ 87
0Y 202 VLLDRGKMD-DLFCBFLGMSWLLPLDFPMTQDFGLNDESSRCYGVKNKNOVIDTEG 260
Db 88 TVVEYETHDKDTDMF-----QVGRSTESPIDFVDTLTISG-SQNTDEA-----OI--TOS 134
0Y 261 TLSHLTYLHLVYDGDHDKMFCEDDQFTIGCVPMPLYKTDNYFPPSLMLTGPFDDELNKL 320
Db 135 TTSRFACRIV-----CDNREPTTAI-----FAAGFSSKNIF 167
0Y 321 FPOKATVRHHLGRYLFHPNTQVMGLVTRYEAYLSHADKIGIOVR-----VEDDEPGP 374
Db 168 LGEKAKAKK-----INDGHMDGLTTN-----GVLVMPHRRGFTTESQPGV 207
0Y 375 FOHV-----MDOISSCTOKELLP-----VD-----TLVRSR---HYNTP 408
Db 208 WRELTSVCGDYVTLRETRSAOGRKLVESEFTVWLQDGLSLIDLCGATLLMTRADGLFHTPTQ 267
0Y 409 KHKAVLYLNSAGVAE-----NLKSMY-----WEYTSJGELIIVH 444
Db 268 KHIEFLROELINARPQCPVGLNTLAFPSINKEVEVEEKOPAY--LSCGHVGHG 320

```

```

1 RESULT 6
2 US-09-843-905A-4
3 : Sequence 4, Application US/09843905A
4 : Patent No. US2002016683A1
5 :
6 : GENERAL INFORMATION:
7 :
8 : APPLICANT: Bird, Timothy A.
9 :
10 : TITLE OF INVENTION: HUMAN PELLINO POLYPEPTIDES
11 :
12 : FILE REFERENCED: 2990-A
13 :
14 : CURRENT APPLICATION NUMBER: US/09/843, 905A
15 :
16 : CURRENT FILING DATE: 2001-04-27
17 :
18 : PRIOR APPLICATION NUMBER: US 60/200,198
19 :
20 : PRIOR FILING DATE: 2000-04-28
21 :
22 : NUMBER OF SEQ ID NOS: 15
23 :
24 : SOFTWARE: Patentln version 3.1
25 :
26 : SEQ ID NO. 4
27 :
28 : LENGTH: 418
29 :
30 : TYPE: PRT
31 :
32 : ORGANISM: Homo sapiens
33 :
34 : US-09-843-905A-4

```

```

Db      18 GELIVLYGNGSLPNGDRGRKRSFALFKRPAKANGVCKPST-----VHIACP--- 63
OY      152 SYKKALKOLDOEHIDODGCKXVWVITSSGLONRLLSLASVFLVALLDRVLVDKGRDM 21
Db      64 QAKAKAISNKDOJIS-----YTLSPRAQTVVEVYTHDS 95
OY      212 D-DLCEPPLGMSWMLPLDFEPMIDOFGLNDESSRCGYMWKNQVIDF---EGTSLHYL 267
Db      96 NIDMF---QIGRSTESPIDEIVTIDVPSQSN-----DIQVSQSTISNRAC 139
OY      268 HLYHYGDHDKKFFCEGQOTFLGKVPMLIVKTDNFEVSLMLIFCFDELLKLPQKATV 322
Db      140 RII-----CERNPPFARI-----YAGFDSKNIPLGEKAAK 172
OY      328 FHHLSRNYLFHPNTQYGVATRYEAYLSHADEKIGIQV---RVFDED--GPFQHV--- 376
Db      173 WK-----TSDGQDGLTTN-----GVLVMPHRNGFTEDSKGIMRELSVC 211
OY      379 -----MDQISCTOCKELLPEYDT-----LVSRHVNTPKRAVL 414
OY      415 VYSLNAGYAENLKSMYWEYPTSTGEIIGVHPDSQGYOQTEKKMHNCKALAEWLLSLTD 478
Db      272 RQEIWAARQ-----CPYGF-----NLAFFSMKRKQYVD 301
OY      475 NLVYASMTFEGYVAOGLGGLKFWLLYPENRTTPDPSGGRAMSEPCFHSPPYDCAKT 534
Db      302 EKQPWVYLVNGCHV---HGYNHW--GNKEERDKDRECPMCRSVGP--YVPLMIGCEAGF 353
OY      535 GIDTG 539
Db      354 YVDAG 358

RESULT 8
US-09-815-242-12623
; Sequence 12623, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/357,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12623
; LENGTH: 876
; TYPE: PRT

```

ORGANISM: Staphylococcus aureus
US-09-815-242-12623

Query Match
Best Local Similarity 20.2%; Score 86.5; DB 10; Length 876;
Matches 95; Conservative 82; Mismatches 151; Indels 143; Gaps 28;

QY 84 GVPFVNTINSKDLGLGSLAGFDEDSCLSR-YOSVHYRKPSYPKSSYLLS-----K 135
DB 213 GVEHLDEISGKRLKEKJONS--KEDALMSKELATINVDSPLEVKLEDTLMHOEOQEK 270
QY 136 LRMTEKLHKRCGPSTESYKALKOLDQ-----EHIDGGECKYVWI 177
DB 271 IELFKKL-----EFKQLADIDOGASVEDAELEKTFELETSPDNVDFTSLKRAVILHF 321
QY 178 SFSGLGNILSLASVFLALLDRVLVDRGKMDLCEPFLGSMILLPDTFPTDQ-- 235
DB 322 ELDC-GNYLRN--NLKFSLFTGEKHIVINADJINN-YAEL--VSW--LENPTKKVY 371
QY 236 FD-----GLNDSRCYGYMKNNOYIDTEGTLSHLYLHVHYGDHDKMFCEG 284
DB 372 YDAKTTYASHRLGIDIONI-SFDMLASYIIDPSRTISDVO-SVYSLYGQS---FYKD 425
QY 285 DQFTIGKVPWLIVKTDNYFVPSLML-----PGFDELNKLFPKATYFHHLGRIYLF 336
DB 426 DVSITGKGGKFKFVEDVLIPIYVASITDAIFYAKPNMDKOLEEY----- 469
QY 337 HPTNQMGLVTRYIE-ATLSHADEKIGIQVRYFDEDPGPFQHVMDQISSCTOKELRPE 394
DB 470 --NOVELLADLEPLAKILSEMEIGIFITVDHLE-----EMEKEIOEK 511
QY 395 VDLVERSRL-----VNTPKHKAV-----LVTSINAGYA---ENLKSMYWEP 434
DB 512 LDVLI-RNIIHVAAGDEINSPKOLGVLFETLOLPVIKRTGTGISTAVDYLEQJGQEHF 570
QY 435 TSTGEIIGVHQPQ-----EGYQOT---EKKMHN--GKALAEMYLISLTD 474
DB 571 I-IDVILEYRLSKLOSTYVEGLQKIVSDQRIHFRFNOGLAOTGRLSVD 620

RESULT 9
US-09-843-905A-2
Sequence 2, Application US/09843905A
Patent No. US2002016863A1
GENERAL INFORMATION:
APPLICANT: Bird, Timothy A.
APPLICANT: Cosman, David J.
TITLE OF INVENTION: HUMAN PELLINO POLYPEPTIDES
FILE REFERENCE: 2990-A
CURRENT APPLICATION NUMBER: US/09/843,905A
CURRENT FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/200,198
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 418
TYPE: PRT
ORGANISM: Mus musculus
US-09-843-905A-2

Query Match
Best Local Similarity 2.9%; Score 86; DB 9; Length 418;
Matches 90; Conservative 47; Mismatches 160; Indels 188; Gaps 22;

QY 99 GGLASGF-----DEDSCLSRYOSVHYRKPSYPKSSYLLSKLRNEKLHKRCGPSTE 151
DB 18 GELIYLVGNGSLPNDGRBKRKSRFALFKRKANGVKPST-----VHIACTP--- 63
QY 152 SYKAKQLDQEHIDGDECKYVWISFGAGNRLSLASVFLVALLDRVLVDRGKMD 211
DB 64 QAAKAIKNDHSHS-----YLSRAQIVVVEVETDHS 95

QY 212 D-DLFCPEPLGNSWMLPLDFPMTDQFDGLNDSRCYGYMKNNOYIDT---EGTLSHLYL 267
DB 96 NTDMF---QIGRSTESPIDFVTDIVPGSQSNS-----DTQSGVSTISRFAC 139
QY 268 HLVDYGHDKMFFCEGDTFICKVPWLIVKTDNYFVPSLMLIPGFDDELNKLFPKATV 327
DB 140 RII-----CERSPPFYARI-----YAAFPDSSKNFIIEGKAAK 172
QY 328 FHHLRILYLFHPTNOVWGLVTRYIEAVLSHADEKIGIQV---RVDEED--PGPFQHV--- 378
DB 173 WK-----TSDGQMDGLTNN-----GVLVHHPNNGFTEDSKPGIMREISVC 212
QY 379 -----MDQISSCTOKREKLPEVDY-----LVRSRHYNTPKHKAVL 414
DB 213 GNVFSLRETRSNQOGRKMW-ELETNOLDGSLIDLCGATLLMRTAEGLSHTPTVAKHLEAL 271
QY 415 VTSLNAGYAENIKSMYWEYPISTGETIGVHQPQSGEYQOQTERKMHNKALAEMLYSLTD 474
DB 272 ROEINARPQ-----CPVGF-----NTLAFPSMKRRDYVD 301
QY 475 NLVTSAMSTFGYVAAGLGSLKFWILYRPNKRTTPRPSGGRAMSMERCFSPFTDCKAKT 534
DB 302 EKQPMVYILNCGHV---HGYNHW--GNKEERDQKDRCEPMCRSVGP--YVPLMLGCEAGF 353
QY 535 GIDTG 539
DB 354 YVDAG 358

RESULT 10
US-09-862-810-4
Sequence 4, Application US/09862810
Patent No. US20020034515A1
GENERAL INFORMATION:
APPLICANT: Mobley, Harry L.T.
APPLICANT: Li, Xin
TITLE OF INVENTION: PROTEUS MIRABILIS-BASED VACCINE
FILE REFERENCE: University of Maryland
CURRENT APPLICATION NUMBER: US/09/862,810
CURRENT FILING DATE: 2001-05-22
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 335
TYPE: PRT
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: PapG fimbrial
US-09-862-810-4

Query Match
Best Local Similarity 2.8%; Score 85.5; DB 10; Length 335;
Matches 46; Conservative 25; Mismatches 70; Indels 75; Gaps 11;

QY 250 MYKNOYIDTEGTLSHLYLHVHDYG-----DHDKMFCEGDTFICKVPWLIVKTDNFFV 304
DB 85 VPKKHVYTNNEG--FNIFLDVQSKYGSWMENENDKDFY-----FFVNGYEW 127
QY 305 PSLW-----LIPGFDELNKLFPKATVFNHL-----GRYLFHPTNOVWGLVTRY 349
DB 128 -DTWTNNGARICFYFGNMKQLNNKF--NDLVFRVLLPYDLPKGYHNF-PVRYIRGIQHHY 183
QY 350 YEAVLSHADEKIGIQVRYFDEDPGPFQHVMDQISSCTOKREKLPEVDY-----E 400
DB 184 YDLMOQH-----YKMPYDOI-----KQLPATNTLMLSFDMVWGGO 218
QY 401 RSRHYNTPKHKAVLYVTSLNAGYAENIKSMYWEYPTS 436
DB 219 PSTQVLNIDHGSIVIDRANGINIASOTLSIYCDVPVS 254

```
RESULT 11
US-09-815-242-13431
; Sequence 13431, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13431
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13431

Query Match      2.8%; Score 85.5; DB 10; Length 502:
Best Local Similarity 18.5%; Pred. No. 8.7; 176; Indels 151; Gaps 29;
Matches 96; Conservative 95; Mismatches 176; Indels 151; Gaps 29;

QY 58 FSMIFHQ-----HPSDS--NRIMGFAEAVLDAGVPPN-----VTNINSDKLLGLLAS 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 36 FQIIFPQAGWVHNNANEIMNSVQSIAGNFISGVKPNQIEAIGITNORETIV----- 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 105 GDEDESCISRYOSVHYRKPSPYKPSYLSIKLRN--YEKLHRCGPGTESYKALK-QL 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 90 -WDKRTGTPYNAIWQS-----RQTAFLAEQLKSOGYVEKFEKTKGLITDAVFSATKRW 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 161 DOEHDG-----DSEC-----KYVWISFSGLN--RIISLAVFLYAL-----LTDRV 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 145 ILDHVEGAOERAKEKELLFGTIDTWLWMLDGAHVTDYSNAARTMYLNIEKMWDEI 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 203 I-----LVDRGDMDLF--CEPFLGMSWLPDLDFEMDQ-----FDGLMOESSRCYG 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 205 LEILNIPKALPEVRSNSBYKTAPEHYGGEVPIISGAGQQAALRQLAFERG---- 260
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 249 VYKQVQVDTESLTHLYLHLVHD-----YGDHDKMFECBQDQFT--GKVP 293
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 261 -VWKN--TYGTGSIIMNTGEEMLSENLLTTIGYINGKVVYVALEGSIFIAISAIO 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 294 WL-----IYKT-----DNVFPVPSLWLPGFDEDLNKLFPKATVYFHHGRLFHPT 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 WLRDGLRWENSPSESEKARSDSHNNDDEVIVPAFTGLGAPYNNQA-----R 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 340 NQVWGLVTR--YYEAYLSHADEKIGIOVRFEDEDPGFQHVWMDQISSCYOKEKLLPEVD- 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 GSVFGL-TRGTSKEDFIKATLQISIAVQVR-----DIIDTMQVDTQTAIQVLKNDG 411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 397 -----TLVRSRHVNTPKKAVLVTSLINAGYAFNLKSMWEYPTSTGE 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 412 GAAMNPLMQFQADILIGIDIAAKNLETTALGAFLAGLSVGYWMDLDEL--KILNFTGE 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 440 IIGVHQPSQEGYQJTEKKMHNG--KALAEWLLSLTDN 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 470 LF--EFSKN--ESRKEQLYKGMKRAKATOVFAFVDD 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-10-028-072-234
; Sequence 234, Application US/10028072
; Publication No. US2003004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gunney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
```


PRIOR FILING DATE:	1997-10-27
PRIOR APPLICATION NUMBER:	60/0633229
PRIOR FILING DATE:	1997-10-27
PRIOR APPLICATION NUMBER:	60/0635500
PRIOR FILING DATE:	1997-10-28
PRIOR APPLICATION NUMBER:	60/0635651
PRIOR FILING DATE:	1997-10-28
PRIOR APPLICATION NUMBER:	60/0637044
PRIOR FILING DATE:	1997-10-29
PRIOR APPLICATION NUMBER:	60/0637333
PRIOR FILING DATE:	1997-10-29
PRIOR APPLICATION NUMBER:	60/0637355
PRIOR FILING DATE:	1997-10-29
PRIOR APPLICATION NUMBER:	60/0637588
PRIOR FILING DATE:	1997-10-29
PRIOR APPLICATION NUMBER:	60/0637555
PRIOR FILING DATE:	1997-10-17
PRIOR APPLICATION NUMBER:	60/0642488
PRIOR FILING DATE:	1997-11-03
PRIOR APPLICATION NUMBER:	60/0648090
PRIOR FILING DATE:	1997-11-07
PRIOR APPLICATION NUMBER:	60/0651866
PRIOR FILING DATE:	1997-11-12
PRIOR APPLICATION NUMBER:	60/0658466
PRIOR FILING DATE:	1997-11-17
PRIOR APPLICATION NUMBER:	60/0663644
PRIOR FILING DATE:	1997-11-21
PRIOR APPLICATION NUMBER:	60/0664533
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/0665111
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/0667700
PRIOR FILING DATE:	1997-11-24
PRIOR APPLICATION NUMBER:	60/0692122
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/0692788
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/0693344
PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/0696944
PRIOR FILING DATE:	1997-12-16
PRIOR APPLICATION NUMBER:	60/0723200
PRIOR FILING DATE:	1998-01-23
PRIOR APPLICATION NUMBER:	60/0736122
PRIOR FILING DATE:	1998-02-04
PRIOR APPLICATION NUMBER:	60/0740866
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/0740922
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/0777921
PRIOR FILING DATE:	1998-03-12
PRIOR APPLICATION NUMBER:	60/0789100
PRIOR FILING DATE:	1998-03-20
PRIOR APPLICATION NUMBER:	60/0792944
PRIOR FILING DATE:	1998-03-25
PRIOR APPLICATION NUMBER:	60/0796633
PRIOR FILING DATE:	1998-02-27
PRIOR APPLICATION NUMBER:	60/0797288
PRIOR FILING DATE:	1998-03-27
PRIOR APPLICATION NUMBER:	60/0801555
PRIOR FILING DATE:	1998-03-31
PRIOR APPLICATION NUMBER:	60/0812033
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/0812299
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/0816585
PRIOR FILING DATE:	1998-04-14
PRIOR APPLICATION NUMBER:	60/0818177
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/0818188
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/0829999
PRIOR FILING DATE:	1998-04-24

PRIOR APPLICATION	NUMBER: 60/08332222
PRIOR FILING DATE: 1998-04-28	
PRIOR APPLICATION	NUMBER: 60/08354545
PRIOR FILING DATE: 1998-04-29	
PRIOR APPLICATION	NUMBER: 60/08460000
PRIOR FILING DATE: 1998-05-07	
PRIOR APPLICATION	NUMBER: 60/08462222
PRIOR FILING DATE: 1998-05-07	
PRIOR APPLICATION	NUMBER: 60/08463737
PRIOR FILING DATE: 1998-05-07	
PRIOR APPLICATION	NUMBER: 60/08514949
PRIOR FILING DATE: 1998-05-12	
PRIOR APPLICATION	NUMBER: 60/08532222
PRIOR FILING DATE: 1998-05-13	
PRIOR APPLICATION	NUMBER: 60/08533838
PRIOR FILING DATE: 1998-05-13	
PRIOR APPLICATION	NUMBER: 60/08533939
PRIOR FILING DATE: 1998-05-13	
PRIOR APPLICATION	NUMBER: 60/08557979
PRIOR FILING DATE: 1998-05-15	
PRIOR APPLICATION	NUMBER: 60/08569999
PRIOR FILING DATE: 1998-05-15	
PRIOR APPLICATION	NUMBER: 60/08570404
PRIOR FILING DATE: 1998-05-15	
PRIOR APPLICATION	NUMBER: 60/08641414
PRIOR FILING DATE: 1998-05-22	
PRIOR APPLICATION	NUMBER: 60/08643030
PRIOR FILING DATE: 1998-05-22	
PRIOR APPLICATION	NUMBER: 60/08710606
PRIOR FILING DATE: 1998-05-28	
PRIOR APPLICATION	NUMBER: 60/08802626
PRIOR FILING DATE: 1998-06-04	
PRIOR APPLICATION	NUMBER: 60/08873030
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION	NUMBER: 60/08874141
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION	NUMBER: 60/08881010
PRIOR FILING DATE: 1998-06-10	
PRIOR APPLICATION	NUMBER: 60/08885555
PRIOR FILING DATE: 19/98-06-11	
PRIOR APPLICATION	NUMBER: 60/08953232
PRIOR FILING DATE: 1998-06-17	
PRIOR APPLICATION	NUMBER: 60/08959999
PRIOR FILING DATE: 1998-06-17	
PRIOR APPLICATION	NUMBER: 60/08990707
PRIOR FILING DATE: 1998-06-18	
PRIOR APPLICATION	NUMBER: 60/08994747
PRIOR FILING DATE: 1998-06-19	
PRIOR APPLICATION	NUMBER: 60/09034949
PRIOR FILING DATE: 1998-06-23	
PRIOR APPLICATION	NUMBER: 60/09042222
PRIOR FILING DATE: 1998-06-24	
PRIOR APPLICATION	NUMBER: 60/09044545
PRIOR FILING DATE: 1998-06-24	
PRIOR APPLICATION	NUMBER: 60/09053838
PRIOR FILING DATE: 1998-06-24	
PRIOR APPLICATION	NUMBER: 60/09086363
PRIOR FILING DATE: 1998-06-26	
PRIOR APPLICATION	NUMBER: 60/09086363
PRIOR FILING DATE: 1998-06-26	
PRIOR APPLICATION	NUMBER: 60/09136060
PRIOR FILING DATE: 1998-07-01	
PRIOR APPLICATION	NUMBER: 60/09151919
PRIOR FILING DATE: 1998-07-02	
PRIOR APPLICATION	NUMBER: 60/09196262
PRIOR FILING DATE: 1998-07-07	

```

Query Match      2.8%  Score 85;  DB 9;  Length 1160;
Best Local Similarity 19.3%  Pred. No. 36;
Matches 106;  Conservative 70;  Mismatches 187;  Indels 186;  Gaps 30;

OY 89 VTN--INSDKLLGGLASGF---DEDSCLSRQSVYKRPSPYKSSY----- 131
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 ITNPFLSD-----IVASSFLKSDNRRIGGGYKKTIVYE--YKDSYDEVAPCWLGF 101
    : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
QY 132 -----LISKLNRYEKLHKRCGPGTESYKK-----ALKOLDQ----- 162
Db 102 LCPVLQAEVGVILLHLKKNFATRPYTIHPHGVFEKESGSLYPDGSSGPKADDSVPPG 161
QY 163 -EHI-----DGGCEKTYVWISFSGLNRIILSLASVFLYALLTDVRLVLD--- 206
Db 162 GSHLYNMTIPBGHAPTADPAC--LTWYIYSHV--DAPRIATIGLIGPLTICKRGALDGS 218
QY 207 --RGKMDM-----LFCPEFLGMSWLP-----LDFPMTDOPGDLNQEISSRCY---GYM 250
Db 219 PQQRQDVHDHDFLLFSVVDENLSMHLNENIATYCSDPASVDKEDETFOESNRMAINGEV 278
QY 251 VKNOVIDTEGTLSHLYL-----HL-----VHDYGDHDKMFCEGDQTFIGK--- 291
Db 279 F-----GNLPFLNMQAOKRVAMHLFGMGNEIDVHTAFHGOMLTTRGHHTDVANIFP 330
QY 292 -----VPMILVKTNDVYFVPSLMLI-----PGFDELNKLFPQKATVPHNLGRLYFHP 338
Db 331 ATFTVTAEMVPM-----EPGTWILISQVNSHFRRDMQALYKVKSC-----SMAPP 374
QY 339 TNOVWGLVTRY-Y-EAY-----LSHADEKIGIOVRFPEDPGP-----FOHVMQISS 384
Db 375 VDLITGKVRQYFIENHLEIOWDYGPMGH--DGSTGKNLR-----EPGISIDKFFQKSSSRIGG 429
QY 385 CTQKEKLLPEVDTLVRSRHYNTPKHKAVALVTSLNAGYAENLKSMYEYPTSTGEIIGVH 444
Db 430 TYMKVRYEAFQDETFQEKMHLEEDRHILGPIVIRAEVGDITQVVFYNNASQPSFM----- 485
QY 445 QPSQEGYQOT-EKKMNGKALAEWYLLSLDNLVTSMSSTFGYVAGLGGLKPIILYRPE 503
Db 486 QPHGVFEKDYEGTYVNDG-----SSYPGLVAKRPEKVTY-----RWTVPPHA 528
QY 504 NRTTPDPSC 512
Db 529 GPTAODPAC 537

RESULT 13
US-10-121-049-234
; Sequence 234, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C17
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 234
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-234

Query Match 2.8%, Score 85; DB 9; Length 1160;
```

```
Best Local Similarity 19.3%; Pred. No. 36;
Matches 106; Conservative 70; Mismatches 187; Indels 186; Gaps 30;

QY 89 VTN--INSDKLIGLLASGE---DEDSCLSRYSQSVHYRKPSYKPSY----- 131
Db 50 ITNQLPDSO-----IYASFLLSKDKNRIGGYTKTKTYKE--YKSDSYTDEVAQPAWLGF 101
QY 132 -----LISKLNRYEKLHKRCGPGTESYKK-----ALKOLDQ----- 162
Db 102 LCPVLQAEVGVILLHLKKNFATRPYTIHPHGVFEKESGSLYPDGSSGPKADDSVPPG 161
QY 163 -EHI-----DGGCEKTYVWISFSGLNRIILSLASVFLYALLTDVRLVLD--- 206
Db 162 GSHLYNMTIPBGHAPTADPAC--LTWYIYSHV--DAPRIATIGLIGPLTICKRGALDGS 218
QY 207 --RGKMDM-----LFCPEFLGMSWLP-----LDFPMTDOPGDLNQEISSRCY---GYM 250
Db 219 PQQRQDVHDHDFLLFSVVDENLSMHLNENIATYCSDPASVDKEDETFOESNRMAINGEV 278
QY 251 VKNOVIDTEGTLSHLYL-----HL-----VHDYGDHDKMFCEGDQTFIGK--- 291
Db 279 F-----GNLPFLNMQAOKRVAMHLFGMGNEIDVHTAFHGOMLTTRGHHTDVANIFP 330
QY 292 -----VPMILVKTNDVYFVPSLMLI-----PGFDELNKLFPQKATVPHNLGRLYFHP 338
Db 331 ATFTVTAEMVPM-----EPGTWILISQVNSHFRRDMQALYKVKSC-----SMAPP 374
QY 339 TNOVWGLVTRY-Y-EAY-----LSHADEKIGIOVRFPEDPGP-----FOHVMQISS 384
Db 375 VDLITGKVRQYFIENHLEIOWDYGPMGH--DGSTGKNLR-----EPGISIDKFFQKSSSRIGG 429
QY 385 CTQKEKLLPEVDTLVRSRHYNTPKHKAVALVTSLNAGYAENLKSMYEYPTSTGEIIGVH 444
Db 430 TYMKVRYEAFQDETFQEKMHLEEDRHILGPIVIRAEVGDITQVVFYNNASQPSFM----- 485
QY 445 QPSQEGYQOT-EKKMNGKALAEWYLLSLDNLVTSMSSTFGYVAGLGGLKPIILYRPE 503
Db 486 QPHGVFEKDYEGTYVNDG-----SSYPGLVAKRPEKVTY-----RWTVPPHA 528
QY 504 NRTTPDPSC 512
Db 529 GPTAODPAC 537

RESULT 14
US-10-123-904-234
; Sequence 234, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C54
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2003, 15:26:28 ; Search time 40 Seconds
(without alignments)
1858.846 Million cell updates/sec

Title: US-10-037-311a-1
Perfect score: 3004
Sequence: 1 MOUNSYRRSSPIRTTGGSS.....GTLPHVRHCEDISWCLKLV 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1687.5	56.2	509	23	ABB90986	Herbicidally activ
2	1550	51.6	535	21	AA650465	Arabidopsis thalia
3	1530	50.9	500	21	AA650466	Arabidopsis thalia
4	1354.5	45.1	426	21	AA650467	Arabidopsis thalia
5	1124.5	37.4	562	23	ABB91595	Herbicidally activ
6	651	21.7	254	21	AA608487	Arabidopsis thalia
7	631	21.0	219	21	AA608488	Arabidopsis thalia
8	457	15.2	146	21	AA608489	Arabidopsis thalia
9	110	3.7	658	19	AA98873	H. pylori GHPD 173
10	103.5	3.4	409	22	ABB52717	Escherichia coli p

11	99.5	3.3	439	22	ABB68433	Drosophila melanog
12	99	3.3	769	23	AAU93188	Arabidopsis transc
13	99	3.3	814	23	AA017586	M catarrhalis MCAL
14	99	3.3	818	22	AB119945	Moraxella catarrha
15	98	3.3	745	22	AA867351	Human neuron proge
16	98	3.3	864	22	AA892894	Human protein sequ
17	97	3.2	446	18	AAW20095	H. pylori cytoplas
18	97	3.2	594	18	AAW21020	H. pylori cytoplas
19	97	3.2	2044	23	ABB47302	Listeria monocytog
20	97	3.2	2506	22	ABG07191	Novel human diagno
21	97	3.2	2506	22	ABG30064	Novel human diagno
22	96.5	3.2	1320	22	ABB66149	Drosophila melano
23	96	3.2	520	23	ABP60956	Mus musculus thior
24	96	3.2	524	23	ABP60954	Mus musculus thior
25	96	3.2	528	23	ABP60955	Mus musculus thior
26	96	3.2	1240	23	ABP29556	Mus musculus thior
27	95	3.2	1802	21	AA818217	Streptococcus poly
28	95	3.2	2932	18	AAW19675	Plasmodium falcipa
29	93	3.1	723	23	ABP40215	ATM mutant 2467del
30	93	3.1	1817	21	AB118301	Staphylococcus epi
31	93	3.1	4134	20	AA731946	Plasmodium falcipa
32	92.5	3.1	407	23	AB806071	Human NS protein s
33	92.5	3.1	933	17	AA895268	Pre-nisin modifica
34	92.5	3.1	933	20	AAV06666	Nisin B of lactoba
35	92	3.1	496	21	AA644216	Arabidopsis thalia
36	92	3.1	808	18	AAW38188	Arabidopsis SCARC
37	92	3.1	808	21	AA828602	Protein encoded by
38	91.5	3.0	1295	22	AB864605	Drosophila melano
39	91	3.0	554	21	AAV91659	Human secreted pro
40	91	3.0	558	22	AA894489	Human protein sequ
41	91	3.0	558	23	AB877728	Murine acid sequen
42	90.5	3.0	419	23	AB807921	Murine pellino-2 p
43	90.5	3.0	452	21	AA607112	Arabidopsis thalia
44	90.5	3.0	502	22	AA838067	Streptococcus pneu
45	90.5	3.0	511	21	AA607111	Arabidopsis thalia

ALIGNMENTS

RESULT 1
ID ABB90986 standard; Protein: 509 AA.
XX
AC ABB90986;
XX
DE 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 197.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN MO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI: 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX

PS Claim 5; SEQ ID NO 197; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
XX Sequence 509 AA;
SQ
Query Match 56.2%; Score 1687.5; DB 23; Length 509;
Best Local Similarity 62.9%; Pred. No. 5e-160;
Matches 326; Conservative 68; Mismatches 93; Indels 31; Gaps 8;
QY 54 VLVAFSMTF-HQHPSPDSNRIMGFAARVLDAGVPNPVTNINS---DKLLGLSLASGFDE 108
Db 3 LLASFNIFFKHQ-----LLGATII--NVGSKDSVKPRDRLLGLGLTADFE 45
QY 109 DSCSRYSQSVHRRKRPKPPSSYLLSKRLNRYEKLHRCGPTESTKALKQDDEHID-- 166
Db 46 DSCSRYSQSVHRRKRPKPPSEYLLSKRLNRYEKLHRCGPTESTKALKQDDEHID-- 105
QY 167 --GDGECKYVWISFSGGNRIILASVFLYALITDRVLTLDVDRGDMDDLCEPPLGMSW 224
Db 106 DSDGECKYIIVWAVYGLGNRIILASVFLYALITDRVLTLDVDRGDMDDLCEPPLGMSW 165
QY 225 LLPLDFPMTDQFDGINQSSRCYGVKRNQYIDTEGLSHLYLHLVHDYGDHKKMFCEG 284
Db 166 LLPLDFPMTDQFDGINQSSRCYGVKRNQYIDTEGLSHLYLHLVHDYGDHKKMFCEG 225
QY 285 DOTFGKXPMILVTDNPFVPSLMLPIGFDDELKLFQKATVFNHILGRYLFHPNQWVG 344
Db 226 DOTFGKXPMILVTDNPFVPSLMLPIGFDDELKLFQKATVFNHILGRYLFHPNQWVG 285
QY 345 LVTRYEAYLSHADEKIGIQVRVEDDEDPFOHVMQDQSSCTOKFKILPEVDTLVERSRH 404
Db 286 WYTSYNAYLSHADEKIGIQVRVEDDEDPFOHVMQDQSSCTOKFKILPEVDTLVERSRH 404
QY 405 VNT---PRHKAIVLTSNAGYAENIKSMWMEYPTSTGEIIGVHOPSGQYQOTEKMHNG 461
Db 345 TNTSNPSTKLKAVLTSNAGYAENIKSMWMEYPTSTGEIIGVHOPSGQYQOTEKMHNG 404
QY 462 KALAEMLYLLTDNLVTGAMSTFGVVAAGLGKLPWILYRPENRTTPPSCGRAMSMPC 521
Db 405 KALAEMLYLLTDNLVTGAMSTFGVVAAGLGKLPWILYRPENRTTPPSCGRAMSMPC 521
QY 522 FHSPPEYDCKAKGTIDTGLVPHVHNCEDI--SWGILKLV 558
Db 465 FLTPPVHGCCEAKKGINTKAIVPEVRHCEDLRHNGKLV 502
RESULT 2
ID AAG50465 standard; Protein: 535 AA.
XX
XX AAG50465:
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63958.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX

PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.

DB 429 DKVHDKALAEYLLSLTDNIIVASSRSTFGYVAYSLGSLKFWLLPLPNDKAPDPCCVR 488
OY 515 AMSMEPCFSPPEYDCKAKT-GIDTGLTPHYRHGCEDISWGTL 557
DB 489 STSMPCFLTPPHGCEPDPAWGTESGKVPFVRYCEDI-WGTL 531

RESULT 3
ID AAG50466 standard; Protein: 500 AA.
XX AAG50466;
AC AAG50466;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63959.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.


```
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161923.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 50.9%; Score 1530; DB 21; Length 500;
Best Local Similarity 58.5%; Pred. No. 3.2e-144;
Matches 299; Conservative 72; Mismatches 118; Indels 22; Gaps 9;
```

```
OY 54 VIVASMFHOHSDNSNRIMGFAEAVIDACGFPPVNTINSKLLGGLASGDFDEDSCLS 113
DB 1 MLTSSNNF-----NNKL--FA-ATINDESETP-----GRDLILGGLTADFDGSCSLS 46
OY 114 RYOSVH-YRKPSYPKPSYLLSKLNRYEKLHKRCGPTGSEYKKAJKQJDOEH----IDOD 168
```

```
DB 47 RYHKTFLYRKPSYPKPSYLLSKLNRYEKLHKRCGPTGATKAYEAKRHLSHDENYAKSD 106
OY 169 GECKTVVWISSEGLNRIILSLASVFLYALLIDRVLVLDGRKMDLFCBEPFLGSMMLPL 228
DB 107 GECRYVWVLADYGLCNRLFLTLASVFLYALLIDRILVLDNRKIDGLCEPFGTSMLLPL 166
OY 229 DEPMTDQFDGLNOESSRCYGVWKNQVVIDTEGTLSHLYLHLVHOXGDHDKMFCGCGDOTF 288
DB 167 DEPLMKYADGYHKGYSCRGYGLMNLNHSINSTISFPPLHYMNLHDSRDSKMFECOKDOSL 226
OY 289 IGVPMILYKTDNVEVPSLMLIPGFDDDELNKLFPKATVFHNLGTYLEPHTQVWGLYVR 348
DB 227 IDKVPMLIFRAVAVFVPSLWMPPTFOTELTKLFPQKEVYFHHGLYLFPRKNQVMDIYTK 286
OY 349 YVEAYLSHADEKIGTQVRFVEDPGRFOHVMQJSSCTQKEKLEPVDLVERSNNV-NT 407
DB 287 YYHDLHSKADERLGIQIRVFRDGGYGVYOHVMDQVISCQDREKLELPLATOESKVNISNI 346
OY 408 PKHKAIVLTVSLNAGYAENLKSWMYEPPTSGELIGVHOPOSQGYOOTEKKHNGKALAE 467
DB 347 PSKAIVLTVSLSPESKLENNFSEKANTGELIKVYOPSGERYQOTDKRVHDQKALAE 406
OY 468 YLLSLTDLNLVTSAMSTFGYVAOGLGLKPMILYRPENRTTDPSCGRAMSMPCFHSPPF 527
DB 407 YLLSLTDLNLVTSAMSTFGYVAOGLGLKPMILYRPENRTTDPSCGRAMSMPCFHSPPF 466
OY 528 YDCKAKT-GIDTGLVPHVRCEDISWGLK 557
DB 467 HCCEPDAMGTEGSKVVPFVRCEDI-WGLKL 496

RESULT 4
AAG50467
ID AAG50467 standard; Protein; 426 AA.
AC AAG50467;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63960.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63960.
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX Arabidopsis thaliana.
OS EP1033405-A2.
XX 06-SEP-2000.
PD 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
```

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137282.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0143390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144083.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145861.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147039.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 45.1%; Score 1354.5; DB 21; Length 426;
Best Local Similarity 60.7%; Pred. No. 1e-126;
Matches 256; Conservative 60; Mismatches 99; Indels 7; Gaps 4;

QY 142 LHKRCGPTESYKALAKOLOEH---IDGDECKYVWISFGHNLILASVFLYL 197
DB 2 LHKRCGPTKAYKEATKHLSDENYNAKSDCEKRYVWLADYGLGNLLTLASFYAL 61
QY 198 LTRVLLVDRGKMDLFCERFLGSMWLLPLDFPMTDOFDGLNDESSRCYGYMKNQYID 257
DB 62 LTRRILVDNRKIDGLLCEPFPPTSMWLLPLDFPLMKYADGYHKGYSRCYGTMLNHSIN 121
QY 258 TEGTSLHYLHYHVGHDHMKFCEGDOTFGKYPMILVKTNDNFVYSMLIPGFDEL 317
DB 122 STSPRPLHYMHLHDSRSDMKFCQKDSLDLKYPMILFRANVYFVSLWNPTEFQEL 181
QY 318 NKLFQKATVFHNLGRYLFHPTNQWGLVTRYEAYLSHADEKIGIYRVFEDDEGPRFH 377
DB 182 TKLFQKQETVHNLGRYLFHPRNQWMDIVTKYHDLKADRLIOQLRVFDQGGYTOH 241
QY 378 VMDQISSTQKELLPEVDLVERSRLH-NPRKRAVLVTSINAGYAENLKSMEYEPYS 436
DB 242 VMDQISCTQREKLPELATQESKVNISNIPKRAVLVTSISPEYSKLEMFSEKRAM 301
QY 437 TGEITGVHQPQEQOQTEKKHNGKALAEWYLLSTLNLVYSAMSTGCVAAOGGLKP 496
DB 302 TGEITGVHQPQEQOQTEKKHNGKALAEWYLLSTLNLVYASSTGCVAAVSLGGLKP 361
QY 497 WILYPRNRTTPRPSGGRMSWEPFCHSPFPFDCAKT-GIDTGLVPHVRHCEDISGL 555
DB 362 WILYPRNRTTPRPSGGRMSWEPFCHSPFPFDCAKT-GIDTGLVPHVRHCEDISGL 555
QY 556 KL 557
DB 421 KL 422

RESULT 5
ABB91595
ID ABB91595 standard; Protein; 562 AA.
AC ABB91595;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 806.
XX
KW Herbicidal; plant; agriculture; herbicide.
OS Arabidopsis thaliana.
XX
PN W0200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX

PR 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB) BAYER AG.
XX
XX Tietjen K, Weidner M;
XX
XX WPI: 2002-269010/31.
XX
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 806; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 562 AA;

Query Match 37.4%; Score 1124.5; DB 23; Length 562;
Best Local Similarity 45.2%; Pred. No. 2e-103;
Matches 242; Conservative 83; Mismatches 163; Indels 47; Gaps 13;

QY 15 TTTGSKSVNFSELQOMYLLSGTMTKLTFTTCLIVESVLAFAFMIFHQHPSDSNRING 74
DB 56 TRTGS-----SELNMMKRPSSLSKMTMGLLAVLAVASVMSLSVLRDPPSD-DVLETT 108
QY 75 FAEARVLDAVPEPNVTNINSKILGL-----LASGDESDCLSRQSVHYHRKP 123
DB 109 EASRVLOSRLH---QATESD---GLSEKKAQLNINLVPSFDEKSCLSRYEASLYRKE 162
QY 124 SPKRSSYLISKLRNVEKHLKRCGPTESYKALAKOLOO-EHIDDD-GECKVYVVISFG 181
DB 163 SPKQSYLDVRLQVREDELHRCGPFTRSYNLTLDKLKSGDSDEVSQCRVIVMLNSNG 222
QY 182 -LGNRLILASVFLAYLLTRVLLVDRGKMDLFCERFLGSMWLLPLDFPMTDOFDGLN 240
DB 223 DLGNMMLSLASFYLLALTLNRLVLELVGDADLCEPRPNTTWLPLPEFPLNSHF---N 279
QY 241 QESSRCYGYMKNQYIDTEGLSHYLHLVHDYGDHMKFCEGDOTFGKYPMILVKTID 300
DB 280 EQS-----LLRNSGNPMVAVRYHVRDSSDQKLFCEDSQVLLLETPMLTKAD 328
QY 301 NYFVPSMLIPGFDELKTLFPQKATVFHNLGRYLFHPTNQWGLVTRYEAYLSHADEK 360
DB 329 SFELPSLFSVSSFKQELMLFPEKDTAFHFLSOLYLFHPTNVWGLITRYNNYLLAKAQR 388
QY 361 IGIOVRVDEDEGPRQHWMDQISCTQKELLPEVDLVER---SRHVTNPRKRAVLVTS 417
DB 389 IGIIYGVSESGNQFOHLIDQILACGTRKRLPEVDK--QRNLPSOVLNKRKSAVPLISS 446
QY 418 INAGYAENLKSMEYEPSTGEITGVHQPQEQOQTEKKHNGKALAEWYLLSTLNL- 476
DB 447 SSPGYFKSIRDYWMENPVTYGEIISVHKPSYKDYKTPRNMBSKRAMEIYLLSCDALV 506
QY 477 VTSAMSTGCVAAOGGLKPLVLYPRNRTTPRPSGGRMSWEPFCHSPFPFDC 531
DB 507 VTGLWSSLVEVAHGLGLKPVWLNRKANGTAHEPYCARARSIEPQSQTLLFGCK 561

RESULT 6
AAG08487
ID AAG08487 standard; Protein; 254 AA.
AC AAG08487;

```
XX 17-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 6045.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-fa2.
XX
XX 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0128845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147436.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
```

PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161952.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	21.7%	Score 651;	DB 21;	Length 254;
Best Local Similarity	53.4%	Pred. No. 1,86-56;		
Matches 132;	Conservative 36;	Mismatches 59;	Indels 20;	Gaps 6;
QY	41	LTTRFTTCLIAFSVLAVASMIFHQHSDBSNRIMGFAEARVLDAGVFPVNTYNSDKLGG	100	
	:	:	:	:
Db	23	ITVFSTLLILSVMLSPSSNNP-----NNK--FA-ATINDESETP-----GRRLLIG	67	
QY	101	LLAGSFDEDSCLSRQSHV-TRKSPRYPPSSLLIKLNRYEKLHRRCGFTSYKALKO	159	
	:	:	:	:
Db	68	LLTTDFDEGSCLSRHKHTFLYRKRPYPKPSSELYLASKLSYEMRNHRCGPGTKAYKEATHN	127	
	:	:	:	:
QY	160	LDQEH----IGDGECKTVVWVIFSEGLNRIILSLASVFLAATPTDRLVLDLNGKIMDDP	215	
	:	:	:	:
Db	128	LSHDENYNAKSDGDCRRVYVWLADYGLGNRLRLTLASVFLYALLTRIILIVDNKRDISDLL	187	
QY	216	CEPFLGMSWLLPPLDFPMTDQFDGLNOESSRCYGIYWKNOVDTBSTLSHYLHLVHYDGD	275	
	:	:	:	:
Db	188	CEPFGTSMWLLPPLDFPMLKYADGYRKGYSRCYGTMLLENHSINSTSPRPLVYMHNLHDSRD	247	
QY	276	HDKMEFC 282		

Db	248	SDKMFC	254
RESULT 7			
AA08488			
ID	AA08488	standard; Protein: 219 AA.	
XX			
AC	AA08488;		
XX			
DT	17-OCT-2000	(first entry)	
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 6046.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
RW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PE	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	06-APR-1999;	99US-0127462.	
PR	08-APR-1999;	99US-0128234.	
PR	16-APR-1999;	99US-0128714.	
PR	19-APR-1999;	99US-0129845.	
PR	21-APR-1999;	99US-0130077.	
PR	23-APR-1999;	99US-0130449.	
PR	28-APR-1999;	99US-0130891.	
PR	30-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	04-MAY-1999;	99US-0132407.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136783.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	16-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139492.	
PR	18-JUN-1999;	99US-0139454.	

[illegible]

PR	17-AUG-1999	9905-0119175
PR	18-AUG-1999	9905-0119426
PR	20-AUG-1999	9905-0119722
PR	20-AUG-1999	9905-0149733
PR	20-AUG-1999	9905-0149293
PR	23-AUG-1999	9905-0149929
PR	23-AUG-1999	9905-0149920
PR	25-AUG-1999	9905-0105666
PR	26-AUG-1999	9905-0150844
PR	27-AUG-1999	9905-0151085
PR	27-AUG-1999	9905-0151060
PR	30-AUG-1999	9905-0151303
PR	31-AUG-1999	9905-0151438
PR	01-SEP-1999	9905-0151930
PR	07-SEP-1999	9905-0152363
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	15-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154779
PR	22-SEP-1999	9905-0155149
PR	23-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0155569
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156456
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157665
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0159293
PR	13-OCT-1999	9905-0159294
PR	14-OCT-1999	9905-0159265
PR	14-OCT-1999	9905-0159329
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159637
PR	18-OCT-1999	9905-0159638
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160747
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814
PR	22-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160988
PR	25-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161922
PR	28-OCT-1999	9905-0161993
PR	29-OCT-1999	9905-0162142

[illegible]


```
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.2%; Score 457; DB 21; Length 146;
Best Local Similarity 58.3%; Pred. No. 2.2e-37;
Matches 84; Conservative 21; Mismatches 35; Indels 4; Gaps 1;
```

```
OY 143 HKRCGPGRSTYKALKOLDQEH---IDGDCCKTYVWISSGCLNRILISLASFVLLAL 198
DB 3 HKRCGPGRSTYKALKOLDQEH---IDGDCCKTYVWISSGCLNRILISLASFVLLAL 198
DB 3 HKRCGPGRSTYKALKOLDQEH---IDGDCCKTYVWISSGCLNRILISLASFVLLAL 198
```

```
OY 199 TDRVLLVDRGKMDLFCFEPFLGMSWLLPDPFMTDQFDGLNOESSRCYGMKNQVIDT 258
DB 63 TDRVLLVDRGKMDLFCFEPFLGMSWLLPDPFMTDQFDGLNOESSRCYGMKNQVIDT 258
OY 259 EGTLSHLYLHVHDYGDHDKMFC 282
DB 123 TSFPHLYMHNHDSRDSKMFEC 146
```

```
RESULT 9
AAM98873
ID AAM98873 standard; Protein; 658 AA.
```

```
AC AAM98873;
AC 31-MAR-1999 (first entry)
DE H. pylori GHP0 1739 protein.
```

```
DE GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
DE peptic ulcer disease.
```

```
OS Helicobacter pylori.
```

```
OS MO9843478-A1.
```

```
PN 08-OCT-1998.
```

```
PF 01-APR-1998; 98WO-0506371.
```

```
PR 29-JUL-1997; 97US-0902615.
```

```
PR 01-APR-1997; 97US-0833457.
```

```
PR 24-JUN-1997; 97US-0881227.
```

```
PA (HUMR-) HUMAN GENOME SCI INC.
```

```
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
```

```
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
```

```
DR WPI; 1998-542293/46.
```

```
DR N-PSDB; AAX14592.
```

```
PS New isolated Helicobacter polynucleotides - used to develop products
```

```
PS for the diagnosis, prevention and treatment of Helicobacter
```

```
PS infections and gastrointestinal diseases
```

```
PS Claim 8; Page 1981-1984; 2054pp; English.
```

```
PS This sequence represents a Helicobacter pylori GHP0 protein of the
```

```
PS invention. The polypeptides can be used for preventing or treating
```

```
PS Helicobacter infections, and gastroduodenal diseases associated with
```

```
PS these infections, including acute, chronic, and atrophic gastritis, and
```

```
PS peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
```

```
PS used for the production of antibodies. The products can also be used for
```

```
PS detection and diagnosis.
```

```
SO Sequence 658 AA;
```

```
Query Match 3.7%; Score 110; DB 19; Length 658;
Best Local Similarity 17.8%; Pred. No. 0.18;
Matches 102; Conservative 79; Mismatches 191; Indels 202; Gaps 23;

OY 37 GTMKITRFTTCLIVFSVLAFAFMIFHOHPSDSNRIMGFARVLDAGVPPNVTNINSK 96
DB 28 GQORLT-FTTLLLA-----LRNHLSEVEILEKFSRKEIYSYLIN--SNKGDG 74
OY 97 ILGLLASEGPEDCLSRYOSVHYRKPSYKPSYLSKLNRVYELKLRGCGPGRSTYKKA 156
DB 75 KF-RLLISESKDTLLSLIDK-NKRKPS--EPYKIVENELFEKWISENTDKLETIFKG 130
OY 157 LKQDQEHIDGDCCKTYVWISSGCLNRILISLASFVLLALDRLVLLV 205
DB 157 LKQDQEHIDGDCCKTYVWISSGCLNRILISLASFVLLALDRLVLLV 205
```



```

Db 131 LKKL-----MIWISLDKGDQDPLFESMNSKDIELTQT---DLIRNYIWE 175
QY 206 DRGKMDLDFCEPLGSMWLLPDPFMTDOF----- 236
Db 176 TEVKKQEDFYNQYKRAMDEERFEQNETLFRNRYRNYLTIKGKIPNEKRYEAFKDYQKK 235
QY 237 ---DGLNOSSRCYGYW-----KNQVIDTEGLSH-----LYLHLVHDYGDH 276
Db 236 GIEIEDLLKDLQKCYGFCOIAFEKEDDKDLNKALSLVLNLEMDVYPLLELYSDYKD- 294
QY 277 DKMFEGSDQTFIGKVPWLVTQDNVFPVSLMLPGF-----DDELNKLFP----- 322
Db 295 -----GVLSKOD--FIPITILIESYICRAVGLGTNSLNKVPSEFTK 335
QY 323 --OKATVFHHLGRLYF-----HPTNQVMGLV-----TRYEATVLSHADERIG 362
Db 336 HICKDEFFKSLKAFVCLTEKQRPNNDEKFKLEFITIDFYKFKNKFLFLELEMFDPKEP 395
QY 363 IQVRVDEDEPGPOHVMQDISCTQKEKLLPEVDLVERSRRV--NTPKRAVLVTSLSA 420
Db 396 V-----DTQKNIENHIMPQTLTPEMQRDLGENFQAIHEKYLHTIGN-----LTLTGYS 444
QY 421 GYAEN-----LKSM-----TWETPST 437
Db 445 KYSNSFOEKRDMEKGFQSSLSKLQSLKDLSEFKEIEKRASDLADWALKLTPTYLE 504
QY 438 GEIGVHPQSOEGYQOTEKKMHNGKALAEWYLS 471
Db 505 AETLEEKPKKKEKKEKEKRYDKLS 538

```

RESULT 10
 ABB52717
 ID ABB52717 standard; Protein; 409 AA.

```

XX AC ABB52717;
XX DT 11-FEB-2002 (first entry)
XX DE Escherichia coli polypeptide SEQ ID NO 835.
XX KM Escherichia coli; B2/D+A-; antinflammatory; antibacterial;
XX KM immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
XX KM systemic infection; non-diarrhoeal infection; septicaemia;
XX KM pyelonephritis; antibiotic resistance.
XX OS Escherichia coli.
XX PN WO200166572-A2.
XX PD 13-SEP-2001.
XX PF 12-MAR-2001; 2001WO-EP03445.
XX PR 10-MAR-2000; 2000FR-0003145.
XX PR 02-FEB-2001; 2001FR-0001449.
XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX DR WPI; 2001-550253/61.
XX PT A library of DNA fragments of Escherichia coli strains for the
XX PT phylogenic determination of a given strain comprises polynucleotides of
XX PT nature B2/D+ A- .
XX PS Example 6; Fig 6; 646pp; English.
XX CC The invention relates to a library of DNA fragments of Escherichia coli
XX CC strains comprising polynucleotides (ABB88577-ABA88729 and ABA89533)
XX CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
XX CC B2/D+A-. The polynucleotides have potential antinflammatory,

```

CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicaemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.

Sequence 409 AA;

Query Match 3.4%; Score 103.5; DB 22; Length 409;

Best Local Similarity 24.1%; Pred. No. 0.36;

Matches 53; Conservative 26; Mismatches 68; Indels 73; Gaps 11;

```

QY 389 EKLPE-----VDLVERSRYHNTPK-----HKAVLVYSLANG--YAENLKMYWEYP 434
Db 193 EELCPSEAAVEDTLFAVSQNDADKRTTSPPEENRDGLVIGEGATLLELEHAKARGA 252
QY 435 TSGETIG-----VHOPSOEGYQOTEKK-----MH 459
Db 253 TTYEIVGFATNCDAAHITQFQRETMQYCMQSLKGLSAQDIIIGYISAGTATRDGMA 312
QY 460 NGKALAEWYLSLTQNL-VTSAMSTFGYVAGLGGLKPMI-----LYRPE-NRTTPD 509
Db 313 ESLATATILY-----GDNPVVSLSKYGFTGLGACGALAEWMSLQMMREGMFAPTLNKRPD 368
QY 510 PSCGRAMSMEPCRHSPPEYDCK-----AKTGIDTGLTV 542
Db 369 PNCG---ALDYIMHARKVDCEFLQSNMFARFGINTSITI 405

```

RESULT 11
 ABB68433
 ID ABB68433 standard; Protein; 439 AA.

```

XX AC ABB68433;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 32091.
XX KM Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL12536.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 32091; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and

```


[illegible][illegible]


```

OY 368 FEDEBPFQHVMDQISCTQKEKLLPEVDTLVERSRHVNTPKHKAALYTSINAGY----- 422
Db 332 -----NYLIRHPSCVNVSKMNAVICSGTYAQYVQTM 363
OY 423 -AENIKSMW--EYPTSTGEITGVHQ---PSQEGYQOTER--KMH-NGKA--LAEMYL 470
Db 364 STONLSMTITRDEYPSNPVLRGINQKAAFQYQPVVMLEKGYTIHWNCPAPRTFTLYLV 423
OY 471 SLTDN-----LYTSAMSTFGYVAGLGGLKPMILYRP 502
Db 424 NENKNDWIRVGLCYPSNTSFOYTFGILORQNGSLSKIEYEP 465

```

Search completed: February 27, 2003, 15:30:40
 Job time : 44 secs

